# SAVING TIME: NEW METHODS AND INSTRUMENTATION FOR RADIO VARIABILITY STUDIES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Laura Grace Spitler

January 2013

SAVING TIME: NEW METHODS AND INSTRUMENTATION FOR RADIO VARIABILITY STUDIES

Laura Grace Spitler, Ph.D.

Cornell University 2013

My thesis describes new instrumentation and signal processing techniques developed for time-domain studies of the radio sky and applies these techniques to a variety of radio astronomical data. Time-domain algorithms were developed for the SERENDIP V survey, a commensal SETI survey operating at the Arecibo Observatory. Along with collaborators at the University of California at Berkeley, I helped develop the high frequency resolution digital FFT spectrometers used to collected the data. No signal with the characteristics of being from an extraterrestrial intelligence was observed. A method for automatically classifying broadband and narrowband signals in raw frequency-time data is presented. It uses both the first and second moments of a spectrum to characterize the how broad or narrowband a signal is. Our applications of this technique to real data show that this algorithm is an effective tool for radio frequency interference excision. A survey for rare, bright radio transients was undertaken with a 3.8 m radio telescope on the roof of the Space Sciences Building on Cornell's campus. This survey involved the end-to-end development of the hardware, software, and data analysis. The data were searched from single, dispersed pulses, but none were found. Multi-frequency observations of the eclipsing, binary white dwarf system J0651 were conducted at the Arecibo Observatory to search for variable emission, both short-duration, "burst-like" and periodic emission. The system has an orbital period of only 12.75 min, and this fast ro-

tation may generate radio emission if the stars are magnetic, but no emission was seen. Five new pulsars, including three Rotating Radio Transients (RRATs), were discovered in a single pulse analysis of 23 months of Pulsar ALFA (PALFA) data collected with the Mock spectrometers. We expanded the existing pipeline to include several new algorithms, including the spectral modulation index and a single pulse rating. In addition to the new discoveries, forty-seven previously known pulsars were redetected. From this work I conclude that considering the time domain is key to fully understanding the radio sky. Time-domain studies require special algorithms and instrumentation and particular attention must be made to managing radio frequency interference.

## BIOGRAPHICAL SKETCH

Laura was born on Thanksgiving Day in 1982 to Charles and Mary Spitler in Des Moines, IA. She was later joined in life by two sisters, Katie and Sam. Laura doesn't remember ever not being interested in astronomy. Her father introduced her to it at a very young age by taking her to Friday evening public lectures at the Drake Observatory in Des Moines. The Drake Observatory was a young kid's dream; full of rooms that smelled funny, a narrow spiral staircase, and haunted by the ghost of the observatory's founder buried in the wall. The lectures were given by a close family friend Herb Schwartz, who continues to give them to this day. To a young kid, the lectures were boring, but she was happy to sit through them if it meant getting a chance to look through the telescopes afterward.

Laura's focus (i.e. obsession) shifted throughout her childhood from ballet to horses to music, but astronomy was one interest that persisted through all. Although she eventually choose to study physic and astronomy in college, music was her primary focus in high school playing double bass in several orchestras and the school jazz band. She was a four-time Iowa All State orchestra member (twice as principal). Music, and in particular the final movement of Beethoven's 9th Symphony, inspired her to study German in high school, a decision that has had repercussions on her trajectory in astronomy. In college she continued to play the bass regularly by taking lessons and playing in the symphony.

After graduating from Theodore Roosevelt High School in 2001, she began studying physics and astronomy at the University of Iowa and quickly got involved in research with Prof. Steven Spangler. Her introduction to radio astronomy came during an REU at the National Radio Astronomy Observatory in Socorro, NM, and she was quickly convinced that radio astronomy is the only

worthy endeavor. During that summer Laura realized she was as interested understanding how radio telescopes work and interferometry as quasars and microquasars. She was resolved to gain experience in radio astronomy instrumentation the following summer.

In what may be the single most influential conversation she has had in her entire life, Michael Rupen, her REU advisor in Socorro, recommended finding a summer internship in radio astronomy instrumentation in Germany. He suggested several people to contact, and to make a long story short, Karl Menten offered her a summer position working with the digital electronics lab (Digitallabor) in the Max-Planck-Institut füer Radioastronomie in Bonn, Germany. The summer after her junior year she made first trip outside the US and to Beethoven's birthplace no less. Despite the fact that they found it "a little unusual for a girl to work in Tecknik", the engineers of the Digitallabor welcomed this strange, American girl into their lab. Her summer was a mix of learning the basics of digital electronics, how to work an oscilloscope in German, soldering, and desperately trying to understand the conversation during morning and afternoon coffee breaks.

Laura graduated with a Bachelor's of Science with honors in May of 2005. She had accepted a position in Cornell University's astronomy department but choose to defer for a year to return to the digital lab in Bonn on a ten-month fellowship from the German Academic Exchange Service (DAAD). She learned some FPGA programming, a bit of astronomy, and a lot of German. Her experience left its marks. She returned a fan of the German national soccer team and a bit more grounded than before.

In August of 2006 Laura arrived in Ithaca. Beyond the usual grad student activities, she was heavily involved in the *Curious? Ask an Astronomy* website,

initially just as a question answer and then later also as website administrator. Laura was part of the group that founded the *Curious?* podcast. Beyond the department Laura played double bass in the Cornell Symphony Orchestra for seven semesters. And beyond campus she took up yoga, braved the hills on her bike, and watched way too much British television. The Ithaca connection that will stick with her the longest though is her cat Alia who was adopted from the Tompkins County SPCA.

After completing her PhD, Laura is starting a post-doc at the Max-Planck-Institut für Radioastronomie in Bonn, Germany.

This work is dedicated to my parents.

## ACKNOWLEDGEMENTS

First I must acknowledge my advisor Jim Cordes for guiding my through graduate school. He was willing to give me room to experiment, fail, and then try again and indulged my interest in instrumentation. I also want to thank my group members, past and present, for all their advice and help: Shami Chatterjee, Julia Deneva, Ryan Shannon, Robert Wharton, and Dusty Madison. Two fantastic undergraduates also helped me in my research. Julia Stone from Barnard College helped enormously with Chapter 3. Josh Burt from Cornell helped with Chapter 6 and contributed significantly with Chapter 2. In addition I want to thank Jill Tarbell, Monica Armstrong, Mary Mulvanerton, Dan Wilcox, and Tom Shannon for their help navigating the details. I would also like to thank Lynn Baker for help with the technical aspects of Appendix A and many interesting conversations over lunch.

I would also like to thank the CASPER community for developing the hardware and software libraries I used to build my instruments. In particular I'd like to acknowledge, in no particular order, Dan Werthimer, Andrew Siemion, Terry Filiba, Mark Wagner, Henry Chen, Billy Mallard, Don Backer, Jason Manley, John Ford, Eric Korpela, and Jeff Cobb. I also received support to travel to several CASPER workshops.

I also wish to thank J. Deneva, R. Minchin, and P. Perillat at Arecibo for their guidance is setting up and conducting the observations reported in Chapter 4, and P. Perillat for helping us verify the issues with our C-band data. We also wish to thank D. Lai and R. Lovelace for several useful discussion. L. Spitler received funds to travel to the Arecibo Observatory for the observations from NAIC.

A hoard of people helped keep me sane during graduate school. The lake house people, Rick Kipphorn, Briony Horgan, Melissa Rice, and Casey Dreier, for beautiful evenings spent next to the lake. The Ranch people, Sharon Gerbode, John Gregoire, and Matt Warkentin, for amazing food and karaoke. Dan Tamayo and Manolis Papastergis for many useful and un-useful conversations. Carl Ferkinhoff for more things than I can list. Also Betsey Adams, Drew Brisbin, Frances Gallart-Marques, Dianne Dietrich, and all of the other astronomy graduate students through the years.

The wish to thank the developer of the numpy.where() function for making my life much, much easier. More seriously though, I thank all of the developers of Python, numpy, scipy, and matplotlib, which were used extensively in this work.

Finally I have to thank my parent's for their years of support and gently pushing me to study astronomy. My sisters for putting up with me through the years. And my cat Alia for many hours spent curled up on my lap purring and napping.

During graduate school I was supported by the Z. Cartter Patten '25 Graduate Fellowship in Astronomy Fund, NASA Space Grant Fellowships, and the NAIC graduate research assistantship.

## TABLE OF CONTENTS

CHAPTER 1

# INTRODUCTION

My thesis work explores new techniques and instrumentation for the study of radio transient, and in particular, large-scale surveys for transients. The study of radio transients improves our understanding of the physical process at work in a variety of astrophysical systems. The variable and transitory nature of these sources requires surveys covering both large fractions of the sky and large amounts of observing time. It is important for a transient detection pipeline to be as robust and automated as possible to analyze the large amount of data that is generated. In this introduction I present important background information that I assume the reader knows when reading the main chapters. First I will discuss some of the scientific background and then describe some signal processing techniques and digital instrumentation in more detail.

## 1.1 Science

### 1.1.1 Transients

Just as each time a new frequency window opened in astronomy, we discovered "unknown unknowns", so to have some of the most important discoveries in astronomy been made when we opened up the time-domain window. Radio pulsars were discovered when Jocelyn Bell Burnell and Antony Hewish observed the radio sky with high time resolution for the first time Hewish et al. (1968). Studies of high redshift Type Ia supernovae first revealed the accelerating expansion of the Universe. Microlensing of extrasolar planets promises to

discover extrasolar systems with properties complementary to other techniques Gaudi (2012). Gamma-ray astronomy has long appreciated transient studies due to the prevalence of gamma-ray bursts.

For the purposes of this work, transient does not imply a one-off event like a supernova or gamma-ray burst. Instead, it is a functional classification for any source observed though individual pulses, rather than persistent or periodic emission. While single pulse detection techniques are required for one-off burst events, they are also useful when applied to variable sources such as pulsars. Particular kinds of pulsar emission are best explored through single pulse studies, namely giant pulses and especially intermittent emitters called Rotating Radio Transients (RRATs). Because of their importance in this work, I will discuss them in more detail in next section.

### 1.1.2 Pulsars

Pulsars are a subset of neutron stars that exhibit regular, pulsed emission. Pulsations have been seen across the entire spectrum, but most frequently in radio or X-ray. This work deals exclusively with radio observations of pulsars, so this discussion is limited to radio pulsars. Stars with masses ranging from six to fifteen $M_\odot$ will become neutron stars at the end of their lives (Lyne & Graham-Smith, 2006). The mass of a neutron star ranges from 1.4 to 2 $M_\odot$, radii of around 10 km, and a density comparable to nuclear density ($7 \times 10^{14}$ g cm$^{-3}$). The lower limit to the mass of a neutron star is set by the Chandrasekhar limit. The upper limit depends on the details of the neutron star equation of state, and the highest measured neutron star mass is 2 $M_\odot$ (Demorest et al., 2010).

During collapse of the progenitor star to neutron star with a radius of $\sim 10$ km, magnetic flux is conserved leaving the new neutron star with a magnetic field of $\sim 10^{12}$ G. The standard model of pulsar emission assumes a mis-aligned rotator where the rotation axis and magnetic axis are misaligned. The result is a system with a time-variable dipole moment that loses energy through the emission of electromagnetic radiation. This radiation comes from rotational energy, slowing the pulsar's rotational period. The first time derivative of the pulsar period ($\dot{P}$) is readily be measured with pulsar timing and is used, along with a pulsar's period, to estimate the magnetic field strength and characteristic age.

In terms of evolution, radio pulsars can be classified into two classes: canonical pulsars and recycled pulsars. Neutron stars enter into a pulsar phase shortly after their formation, and what fraction of young neutron stars are also pulsars is unknown. These pulsars are called canonical pulsars and comprise the majority of the pulsar population. Eventually a pulsar's rotation slows enough that the pulse emission mechanism shuts off. If a pulsar is in a binary system, it can experience a second pulsar phase after its rotational period is spun up through accretion of mass from its companion. These recycled pulsars, also called millisecond pulsars, have the shortest periods of all pulsars ($1.5 \text{ ms} \lesssim P \lesssim 30 \text{ ms}$) and are particularly stable rotators (Lorimer, 2008).

Pulsars are not only laboratories for extreme physics, but also useful as tools to study a wide range of phenomenon. Pulsars are precise astrophysical clocks scattered through the Universe and allow us to perform experiments that might otherwise only be *Gedankenexperiments*. The properties of the ionized interstellar medium can be determined using the dispersion and scattering of pulsar pulses (Cordes & Lazio, 2002). The first planet discovered outside of our solar system

was in orbit around a pulsar (Wolszczan & Frail, 1992), proving definitively that extra-solar planets exist. Binary systems containing a pulsar and a second compact object (i.e. white dwarf or neutron star), provide some of the most stringent tests of Einstein's general theory relativity. Notably, the observed orbital decay of the pulsar–neutron star system B1913+16 closely matched the predicted decay due to emission of gravitational radiation (Hulse & Taylor, 1975; Weisberg & Taylor, 2005), proving the first indirect evidence for the existence of gravitational radiation. Several collaborations around the world are using ensembles of millisecond pulsars to search for a background of low frequency gravitational radiation (Demorest et al., 2012; Yardley et al., 2011; van Haasteren et al., 2011).

For further reading on pulsars, there are two standard textbooks on the subject: *Handbook of Pulsar Astronomy* (Lorimer & Kramer, 2004) and *Pulsar Astronomy* (Lyne & Graham-Smith, 2006).

## 1.2 Techniques

### 1.2.1 Dispersion

A radio wave traveling through a diffuse plasma has a propagation velocity that depends on the frequency of the electromagnetic wave ($\nu$). This effect can be described as a frequency-dependent index of refraction

$$n = \sqrt{1 - \frac{\nu_p^2}{\nu^2}} \tag{1.1}$$

where $\nu_p^2$ is the electron plasma frequency. Mathematically, the electron plasma frequency is

$$\nu_p^2 = \frac{2Ne^2}{m_e} \tag{1.2}$$

where $m_e$ is the electron mass, $e$ is the fundamental electron charge, and N is the number density of electrons. Conceptually, the plasma frequency represents an intrinsic time scale that exists in a plasma of a certain density. Given a sea of electrons and protons that is electrically neutral, displacing some electrons relative to the protons will result in them oscillating in response to a Coulomb restoring force with a frequency equal to the plasma frequency. Equation 1.1 also tells us that higher radio frequencies travel faster than lower frequencies (i.e. as $\nu \to \infty$, $n \to 1$). This frequency-dependent arrive time is called frequency dispersion.

A radio wave can only propagate through a medium characterized by a plasma frequency $\nu_p$ if $\nu > \nu_p$. Mathematically, if this condition does not hold, Equation 1.1 gives an imaginary index of refraction, which implies absorption. Conceptually, a radio wave displaces the electrons in a plasma, and if the time scale for one oscillation of the radio wave is shorter than the time scale for one oscillation due to the restoring force, the radio wave can propagate freely. But if the time scale for one oscillation of the radio wave is longer than the restoration time of the electrons, the electromagnetic wave looses its coherence and is absorbed.

The frequency-dependent index of refraction has important observational implications. In a short duration radio pulse, all frequency components leave the source at the same time. But due to the frequency-dependent propagation speed, the higher frequency components arrive before the lower frequency com-

ponents. The amount of frequency dispersion, and therefore the difference in arrival times, depends on the number of electrons along the line of sight between the source and the observer. The difference in arrival times is given by

$$t_1 - t_2 = \frac{e^2}{2\pi c m_e}\left(\frac{1}{\nu_1^2} - \frac{1}{\nu_2^2}\right)\int_0^D N(d)\ \mathrm{d}d \tag{1.3}$$

where $t_1$ and $t_2$ are the arrival times of the pulse at frequencies $\nu_1$ and $\nu_2$ respectively, $c$ is the speed of light in a vacuum, $N(d)$ is the number density of electrons as a function of distance $d$, and the integral is the column density of electrons between the observer and a source at a distance $D$.

In practice we do not know $N(d)$ or D. Instead, measurements of the difference in arrival times ($t_1 = t_2$) and knowledge of $\nu_1$ and $\nu_2$ give us the value of the integral. In the pulsar community it is customary to discuss the amount of frequency dispersion in terms of the a quantify called the dispersion measure, which is defined as

$$\mathrm{DM} = \int_0^\infty \left(\frac{\mathrm{N}}{\mathrm{cm}^{-3}}\right)\mathrm{d}\left(\frac{\mathrm{d}}{\mathrm{pc}}\right). \tag{1.4}$$

Dispersion measure therefore carries the practical units of $\mathrm{cm}^{-3}$ pc. For completeness, we rewrite Equation 1.3 in practical units:

$$t_1 - t_2 = 4.16 \times 10^6\ \mathrm{DM}\left[\frac{1}{\nu_{1,\mathrm{GHz}}^2} - \frac{1}{\nu_{2,\mathrm{GHz}}^2}\right]\mathrm{ms} \tag{1.5}$$

where $\nu_{1,\mathrm{GHz}}$ and $\nu_{2,\mathrm{GHz}}$ are radio frequency in GHz. A full derivation of these dispersion relations can be found in Rohlfs & Wilson (2004).

In practice the Galactic electron density varies with both distance and line-of-sight. Two sources at the same line of sight but different distances will have different dispersion measures (with the more distant source having a higher dispersion measure). Also, two sources at the same distance but at different

| $l$ | $b$ | $D$ (kpc) | DM (pc cm$^{-3}$) |
|---|---|---|---|
| 0 | 0 | 1 | 35 |
| 0 | 0 | 8.5 | 1577 |
| 45 | 0 | 5 | 175 |
| 45 | 0 | 10 | 491 |
| 90 | 0 | 10 | 292 |
| 180 | 0 | 10 | 188 |
| 0 | 90 | 10 | 35 |
| 0 | 180 | 10 | 20 |

Table 1.1: Example values for dispersion measure for various line-of-sight based on the NE2001 model (Cordes & Lazio, 2002). The first two columns are Galactic longitude ($l$) and Galactic latitude ($b$). The third column is distance from the Earth in kpc, and the fourth column is the dispersion measure.

lines-of-sight also will have different dispersion measures. Mapping dispersion measure is one way study the density distribution of Galactic electrons. Or conversely, given a model of the electron density distribution, one can use a measurement of the dispersion measure to estimate the distance to a source.

Dispersion measure is highest in the Galactic plane with the highest value in the direction of the Galactic center. A typical value for the galactic electron density is $n_e = 0.025$ cm$^{-3}$ but can vary by a factor of two depending on the line-of-sight. A source a distance of 1 kpc from the Earth located in the Galactic plane will have a dispersion measure of about 25 pc cm$^{-3}$. In Table 1.1 we list several values of DM for various lines-of-sight and distances in the Galaxy based on the electron density model of Cordes & Lazio (2002).

### 1.2.2 Dedispersion

For studies of short-duration radio pulses, most notably pulsars, the effects of dispersion need to be compensated for in a process called dedispersion. The are two methods of dedispersion: coherent and incoherent. Coherent dedispersion allows the effects of dispersion to be removed completely by adjusting the phase of the time-domain, voltage signal. This method is not used in any of the chapters, so we do not discuss it further. Instead, this work uses incoherent dedispersion.

Incoherent dedispersion is used with intensity-like frequency-time data generated with digital spectrometers like the the ones described in Section 1.3. These data are a series of spectra with some total bandwidth $B$ and frequency resolution $\Delta\nu$ and are separated in time by a fixed interval $\Delta t_s$. Incoherent dedispersion shifts frequency channels in time according to Equation 1.5. Incoherent dedispersion is universally used in surveys for new sources where the dispersion measure of the source is not known *a priori*. Instead a series of trial dispersion measures is applied in turn with the correct value yielding the highest SNR.

Figure 1.1 illustrates the importance of dedispersion with simulated data. Simulated time-frequency data was generated containing Gaussian random noise and a single dispersed pulse with a dispersion measure of DM = 100 pc cm$^{-3}$ and Gaussian width of eight samples. Each of the three frames shows the same simulated data after applying three different levels of dedispersion. The main panels in each frame show the simulated time-frequency data, and the bottom panel shows the time series in units of signal-to-noise ratio generated by averaging over the frequency channels. In pulsar work, it is standard to sum

the frequency channels to boost the SNR of broadband pulses. The top frame shows the data with no dedispersion applied. The dispersion sweep across the simulated 100 MHz bandwidth is about 30 ms. The pulse is too dispersed to be detectable in the time series. (This pulse could be detected using more sophisticated techniques such as a running variance test.) In the middle panel the data have been incorrectly dedispersed with DM = 80 pc cm$^{-3}$, yielding a DM error of $\Delta$DM = 20pc cm$^{-3}$. The total dispersion sweep has been reduced to about 5 ms. A pulse is now visible in the time series but only at SNR $\approx$ 3. Generally detection pipelines apply a minimum SNR detection threshold of 5 or 6, so this pulse would also not be detected. In the bottom frame, the correct dedispersion value was applied. The pulse is now vertical in the time-frequency data, and a bright pulse with a SNR = 25 is visible in the time series. Clearly it is important to correct for dispersion.

## 1.3 Instrumentation

All three of the instruments I developed during my PhD used general-purpose hardware boards and software libraries developed by CASPER[1], the Collaboration for Astronomy Signal Processing and Electronics Research. CASPER brings together people from many institutions for one goal: expedite the development of new digital astronomy systems by using common tools and sharing experience. It is common in science to have a large group of people working together to achieve a common scientific goal. CASPER takes this approach and applies it to hardware and software. Instead of each group or observatory designing and building its own custom instrument, CASPER pushes people to use a com-

[1] https://casper.berkeley.edu/

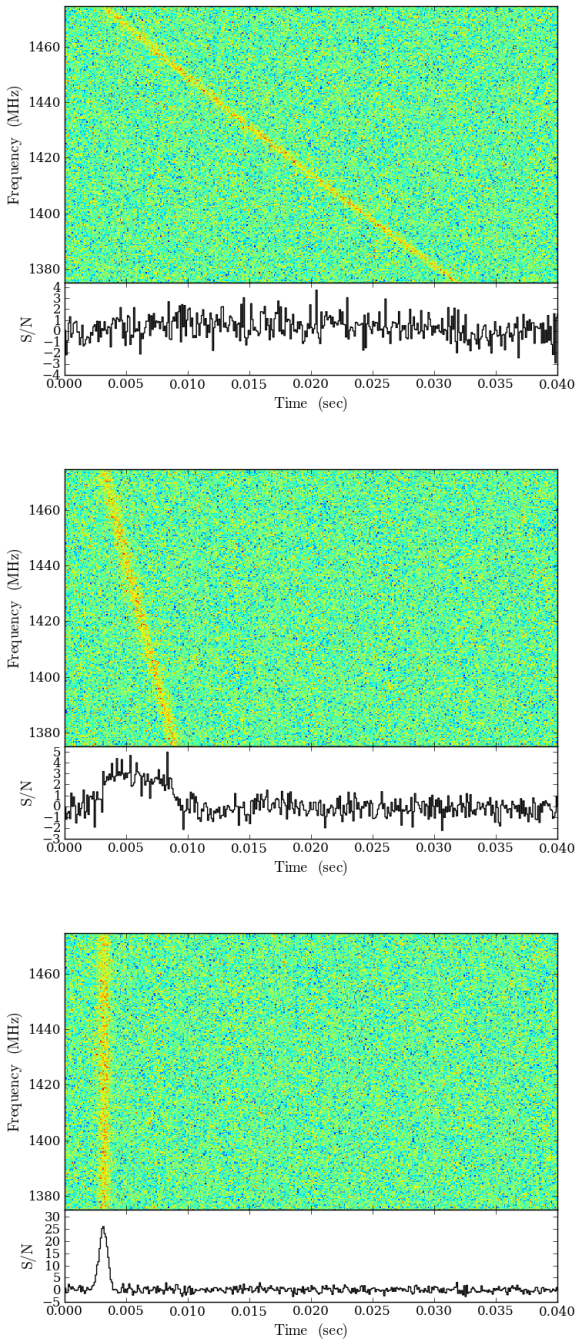


Figure 1.1: Illustration of the importance of dedispersion.

mon set of hardware and software libraries, so any instrument developed by one group can be shared with another group with similar goals.

CASPER was conceived of by Dan Werthimer and Don Backer at the University of California, Berkeley. The technological inspiration for CASPER was the FPGA (Field Programmable Gate Array). An FPGA is a general-purpose integrated circuit with a greater capacity for realtime digital signal processing than the more familiar CPU. An FPGA is a "Gate Array" in that it is comprised of a large number of logic gates. Digital computation is, at its core, simply binary logic. The connections in an FPGA are specified by the user according to her needs, i.e. it is programmable "in the field". This flexibility of allowing the user to specify the connections results in a versatile, fast signal processor. It is also clear that it is useful as the foundation for general-purpose hardware; I want an FPGA to calculate FFTs, while others may need correlators or digital mixers.

All of the instruments I developed were digital spectrometers. The goal of a spectrometer is to convert an analog, time domain radio signal from a telescope into a digital frequency spectrum. This is accomplished in two main steps. First the analog signal is digitized using a type of integrated circuit called an analog-to-digital converter (ADC). The number of bits into which the analog signal is digitized varies with chip but is generally between two and twelve. The higher the number of bits, the larger the dynamic range of the signal, which is an important consideration for environments with a lot of RFI (radio frequency interference).

The second main stage in a spectrometer is the conversion from a time-domain signal to a frequency-domain signal using a Fourier Transform. Because our signal is digitized, we need a discrete Fourier Transform (DFT). An

efficient implementation of a DFT called the Fast Fourier Transform (FFT) is universally used. The primary property of a FFT is the length of the transform, which corresponds to the number of time samples needed for the calculation and the resulting number of frequency channels.

A spectrometer is defined by four properties: total frequency bandwidth, the frequency resolution, the number of frequency channels, and the time resolution (or "dump time"). The total frequency bandwidth is the difference between the highest and lowest frequency in a spectrum and depends on the rate at which the analog data is digitized. The faster the signal is digitized, the wider the total frequency bandwidth. Each frequency channel in the spectrum has finite frequency width, i.e. the frequency resolution, which is set by the total bandwidth and number of frequency channels. The time resolution is the amount of time between new spectra. The minimum time resolution is equal to the total length of time that goes into a single transform, but in practice, a spectrometer will accumulate many spectra before presenting the spectrum.

## 1.4 Chapter Summaries

In Chapter 2 we describe the first results from the SERENDIP V survey operating at the Arecibo Observatory. SERENDIP V is a SETI experiment that operates commensually on observations made with the Arecibo L-band Feed Array. Our analysis here presents results from data taken in 2010 and no candidate signal from an extraterrestrial intelligence was found.

A new signal processing technique used to better identify radio frequency interference in high time resolution radio astronomy data is described in Chap-

ter 3. In general our technique uses multiple moments of a spectrum to better characterize a signal, but we focus on a statistic, the spectral modulation index, that uses the first and second moments to measure how narrow or broadband a signal is. This method is used in Chapters 4, 5, and 6.

Observations of the binary, detached, eclipsing white dwarf system, J0651, were undertaken at the Arecibo Observatory and described in Chapter 4. The system has an orbital period of only 12.75 min . This rapid rotation, along with the possibility of the stars being magnetic, suggests the binary may also be a radio emitter. Our multi-frequency observations revealed no short-duration transient emission or emission modulated at the orbital period.

In Chapter 5 I discuss the results from a survey for bright, rare radio transients conducted with a small radio telescope from the roof of the Space Sciences Building. No transients were seen in over 100 days of observing. We present an event rate limit for this survey and compare it to other current and future transient surveys.

Results from new single pulse analyses of twenty-three months of data from the Pulsar ALFA (PALFA) survey are presented in Chapter 6. The PALFA survey searches the Galactic plane visible from the Arecibo Observatory for new pulsars. In my analysis five new pulsars were discovered and 47 previously known pulsars were redetected just with their single pulses alone.

Upgrades to the radio telescope on the roof the of Space Sciences used in Chapter 5 are described in Appendix A.

CHAPTER 2

# SERENDIP V: A COMMENSAL SETI SURVEY AT THE ARECIBO OBSERVATORY

We present initial results from the SERENDIP V survey operating at the Arecibo Observatory. SERENDIP V is a SETI (Search for Extraterrestrial Intelligence) experiment that operates commensually with the Arecibo L-band Feed Array receiver, a seven-pixel, dual polarization feed array used by many large-scale surveys. The data are recorded with the SERENDIP V.v spectrometer. SERENDIP V.v processes 200 MHz of bandwidth at a frequency resolution of 1.5 Hz and time resoluton of 0.67 s. To reduce the data rate, the spectrometer only records spectral channels that have a signal-to-noise ratio above an adjustable threshold. Different surveys operate in different observing modes (e.g. drift scan, targeted). We chose to focus this analysis on data collected during targeted observations, because they have the greatest potential for identifying good candidates with few reobservations of the same sky position. In this work we present a new analysis technique based on time-domain clustering and present results for the data collected in 2010.[1]

## 2.1 Introduction

The roots of radio SETI go back to the early days of radio astronomy itself when Cocconi & Morrison (1959) suggested searching for signs of extraterrestrial intelligence at frequencies on the order of ~ GHz due to the lack of absorption by the interstellar medium and low sky brightness temperatures. Furthermore,

[1]L. G. Spitler, J. M. Cordes, J. B. Burt (Cornell University), A. Siemion, D. Werthimer, J. Cobb, T. Filiba, E. Korpela, W. Mallard, M. Wagner (University of California, Berkeley), H. Chen (University of California, Los Angeles)

they singled out the 1420 MHz spectral line of neutral hydrogen as a universal, "magic" frequency. An extraterrestrial intelligence wanting to signal its presence may transmit a beacon at a "magic" frequency expecting all radio astronomers to be observing at this frequency. Project Ozma, the first radio SETI experiment, was conducted in 1961 at Green Bank, WV using a search strategy now standard in radio SETI surveys; namely, an observation frequency centered on 1420 MHz and high frequency resolution ($\Delta\nu$=100 Hz) to search for spectral lines (Drake, 1961). A more down-to-earth reason is the wide-spread availability of high quality feeds (and more recently feed arrays) driven by the astrophysical importance of neutral hydrogen in the Universe.

The high spectral resolution is chosen to discriminate artificial signals from natural ones. Doppler motions of astrophysical emitters result in a minimal fractional bandwidth of $\Delta\nu/\nu \approx 10^{-5}$. The narrowest known astrophysical emission are masers with frequency widths on the order of 100 Hz to 1 kHz. Spectral broadening due to interstellar scattering is negligible at radio frequencies around 1 GHz (Cordes & Lazio, 1991). If a celestial signal with a fractional bandwidth significantly less than $10^{-5}$ were to be discovered, it would be strong evidence for an artificially-generated signal from extraterrestrial technology.

The collection of experiments collectively known as SERENDIP (Search for Extraterrestiral Radio Emissions from Nearby Developed Intelligent Populations) is defined by two main characteristics: observations are commensal and a threshold in signal-to-noise ratio (SNR) is applied in realtime. Commensal observations surrender control over telescope pointing and frequency selection in exchange for substantially more observation time than is practical for targeted surveys of nearby-stars thought hospitable to life (Backus & Project Phoenix

Team, 2002). Furthermore, the large amount of observation time presents a data storage and analysis challenge. The SERENDIP experiments reduce the data rate by only recording information for a frequency channel with large enough SNR. The primary data product is sparsely sampled spectra consisting of "hits".

The first SERENDIP experiment was conducted at the Hat Creek Radio Observatory from 1979 to 1982 and then at the Goldstone Deep Space Network antenna in 1981 and 1982 (Bowyer et al., 1983, 1995). It deployed a 100 channel spectrometer with 500 Hz spectral resolution across 100 kHz of instantaneous bandwidth. The instrument stepped through the 20 MHz receiver bandwidth, which was centered on a RF frequency chosen by the observer. SERENDIP II operated at the NRAO 300-ft telescope in Green Bank, WV from 1986 until the telescope's collapse in 1988 (Bowyer et al., 1993). It had a spectral resolution of 1 Hz and scanned a bandwidth of 1.2 MHz in 20 kHz steps. The majority of the observations were conducted near the 1420 MHz neutral hydrogen line and 1667 OH line. In this experiment 3.2 trillion spectral bins were analyzed and millions of events recorded for offline processing. Of these 126 candidates were classified as candidate signals and re-observed with no success.

The third generation SERENDIP experiment was the first to operate at the Arecibo Observatory, and unlike its predecessors, observed at 430 MHz (Bowyer et al., 1995). The spectrometer had a spectral resolution of 0.6 Hz and 2.5 MHz of instantaneous bandwidth. The entire 12 MHz IF was stepped through in 2.4 MHz steps every 8.5 seconds. The survey operated from 1992 to 1996 and observed 93 % of the Arecibo sky at least once and 44 % of the sky at least 5 times (Bowyer et al., 2000).

The SERENDIP IV spectrometer was installed at Arecibo in June of 1997

(Cobb et al., 2000) and observed with the 430 MHz line feed until September of 1998 when observations began using a dedicated L-band feed (Werthimer et al., 2000). The spectrometer processed 100 MHz of instantaneous bandwidth at a spectral resolution of 0.6 MHz. This was accomplished by breaking the 100 MHz band into 2.5 MHz subbands, which were each processed by a dedicated FFT board. This modular, scaleable approach to the hardware allowed versions to be deployed on other radio telescopes. The Southern SERENDIP spectrometer was installed on the Parkes radio telescope in Australia (Stootman et al., 2000). It was a dual spectrometer system that processed 17.6 MHz of bandwidth at a spectral resolution of 0.6 Hz on two of the thirteen beams of the L-band Multipixel Receiver (Staveley-Smith et al., 1996). A scaled-down version of the SERENDIP IV spectrometer was also installed on the Medicina radio telescope in Italy in 1998 (Montebugnoli et al., 2000).

The SERENDIP experiment adopts a wide but shallow approach to SETI, whereby a wide bandwidth is observed but only the signals with a large SNR are analyzed. This is complementary to other experiments such as SETI@home (Korpela et al., 2004), which adopts a narrow but deep approach. SETI@home only observes a 2.5 MHz band but analyzes the baseband signal directly, which allows for much deeper exploration of the data in terms of SNR and signal types. A wider bandwidth increases the likelihood of an ETI signal falling in the observing band, but only analyzing the brightest signals limits the parameter space to brighter and closer transmitters. The large amount of data generated by the SETI@home experiment is analyzed by a volunteer distributed computing network (Korpela, 2012).

In this paper we describe the fifth generation SERENDIP experiment and

present results from the data collected in 2010. The SERENDIP V.v spectrometer was deployed at Arecibo in June of 2009 and observes commensally on the seven-beam Arecibo L-band Feed Array (ALFA)[2]. It observes 200 MHz of instantaneous bandwidth with a spectral resolution of 1.5 Hz and a 0.67 s integration time.

In Section 2.2 we describe how the a candidate signal from an extraterrestrial intelligence (ETI) may vary in time and frequency. In Section 2.3 we describe the survey in detail, and in Section 2.4 we describe the unique characteristics of our data. The time-domain clustering algorithms are described in Section 2.5 and the results when these are applied to the data given in Section 2.6. Discussion of these results and final conclusions are given in Sections 2.7 and 2.8.

## 2.2 Characteristics of an ETI signal

### 2.2.1 Signal variability

A steady, narrowband RF signal transmitted at a fixed frequency by an extraterrestrial intelligence (ETI) will appear to us to be variable in both time and frequency. The rotation of the planet will cause an otherwise steady transmitter, or ensemble of transmitters, to vary in intensity. Uni-directional transmitters will appear to turn on and off as they rise and set (Sullivan et al., 1978). A steady, directional transmitter will appear to us as a Gaussian-shaped pulse as the beam crosses our line of sight. As the planet orbits its sun, the diurnal variations will drift periodically in time over the course of one orbit. If the ETI lives on a large

[2] http://www.naic.edu/alfa/

moon orbiting a gas giant, a third sinusoidal change in signal will be imposed by the moon's orbit around the planet.

If the planet is sufficiently distant, diffractive interstellar scintillations (DISS) will modulate the transmitter's intensity. The intensity scintillations for a planet at a distance of 1 kpc will vary on a time scale of ~3000 s and a frequency scale of ~1 MHz for an observing frequency of 1 GHz (Cordes & Lazio, 1991). For ultra-high resolution SETI spectrometers, the frequency and time resolution are both much smaller than the scintillation time and frequency scales for even the most extreme regions of the galaxy, and the intensity variations of the ETI signal are exponentially distributed with a mean (and standard deviation) equal to the intrinsic intensity. If the observations span many scintillation time scales, the transmitter's intensity is greater than the mean for $\sim 1/e$ of the time and less than the mean for the remainder of the time. This has important implications for the confirmation of candidate ETI signals (Cordes et al., 1997). If a real signal is detected while the transmitter's intensity is greater than the mean, the probability is high that during a later observation the intensity lower and not re-detected at the same SNR threshold.

The orbital and rotational motion of the planet as described above will also impart a Doppler shift on a narrowband signal. The maximum Doppler shift for a planetary system viewed edge on is

$$\Delta\nu_{\max} = \nu_0 \frac{V_0}{c} \tag{2.1}$$

where $\nu_0$ is the intrinsic frequency of the transmitter, $\Delta\nu_{\max} = \nu - \nu_0$, $V_0$ is the relative line-of-sight velocity of the exoplanet and the observer's reference frame, and $c$ is the speed of light. Generally the topocentric frequency as measured by the telescope is converted into a barycentric frequency. Over the course of half

an orbit of the exoplanet, the frequency of a narrowband transmitter would appear to us to shift from $\nu_0 - \Delta\nu_{\max}$ to $\nu_0 + \Delta\nu_{\max}$. Surveys which observe the same position in the sky multiple times or follow-up observations of a candidate signal must correlate signals over a range of frequencies comparable to $\Delta\nu_{\max}$.

The rate of change in the frequency of a signal (i.e. the drift rate $\Delta\dot{\nu}_{\max}$) of a narrowband signal is given by

$$\Delta\dot{\nu}_{\max} = 2\pi\nu_0 \frac{V_0}{c} \frac{1}{P} \tag{2.2}$$

where P is the orbital or rotation period of the planet. Table 2.1 gives characteristic velocities, Doppler shifts, and Doppler drift rates for an observing frequency of 1.4 GHz using the Earth's orbital and equatorial rotational velocity as representative of an exoplanet and Europa's orbital motion as representative of an exomoon.

The drift rate may be important for long observations of the same position in the sky. For the ultra-high resolutions used in SETI spectrometers, it is possible for a signal to drift across several frequency bins in a single observation. The time required for a signal to drift across a single frequency channel ($T$) is given by

$$\Delta\dot{\nu}_{\max} T = \Delta\nu. \tag{2.3}$$

For the SERENDIP V.v spectrometer the frequency drift timescale for the Earth's orbital motion is 50 s, Earth's rotation is 5.7 s, and Europa's orbit is 1.9 s.

| | $V_0$ | $\Delta\nu_{\max}$ | $\Delta\dot{\nu}_{\max}$ |
|---|---|---|---|
| | (km/s) | (kHz) | (Hz/s) |
| Earth's orbit | 30 | 140 | 0.03 |
| Europa's orbit | 14 | 65 | 1.3 |
| Earth's rotation | 0.5 | 2.3 | 0.17 |

Table 2.1: Typical velocities, Doppler shifts, and Doppler drift rates for typical planetary motions.

### 2.2.2 Volume Sampled

In this section we estimate the maximum distance a narrowband signal could be detected with the Arecibo Observatory using the ALFA receiver and SERENDIP V.v spectrometer. We assume a narrowband transmitter radiates a power, $L_\nu$, into a bandwidth of $\Delta\nu = 1$ Hz and a beaming factor of $\Omega_t$ where $\Omega_t = (\pi/4)\theta^2_{\text{HPBW}}$ and $\theta_{\text{HPBW}}$ is the half-power beam-width of the beam. The maximum detectable distance is given by

$$D_{\max} = 0.1\text{pc}\left[\left(\frac{4\pi}{\Omega_t}\right)\left(\frac{L_\nu}{1\text{ MW}}\right)\left(\frac{10\text{ Jy}}{S_{\min}}\right)\right]^{1/2} \tag{2.4}$$

where $S_{\min}$ is the minimum detectable flux density given by the radiometer equation. Figure 2.1 shows the maximum detectable distance as a function of transmitter power for three transmitter beaming angles $\theta_{\text{HPBW}} = 1$ arcmin, 3.5 arcmin, 1 deg, and an isotropic transmitter. Assuming a system equivalent flux density for ALFA of 3 Jy, $\sqrt{\Delta\nu\Delta t_s} = 1$, and a minimum SNR threshold of 10, we obtain $S_{\min} = 30$ Jy. It is also assumed that the frequency of the narrowband signal falls in the center of a frequency bin.

One of strongest radio transmissions humans produce is the Arecibo S-band planetary radar, which transmits 1 MW of power into a 1 Hz bandwidth. Arecibo could detect an extraterrestrial Arecibo pointed directly at us out to a couple kiloparsecs.

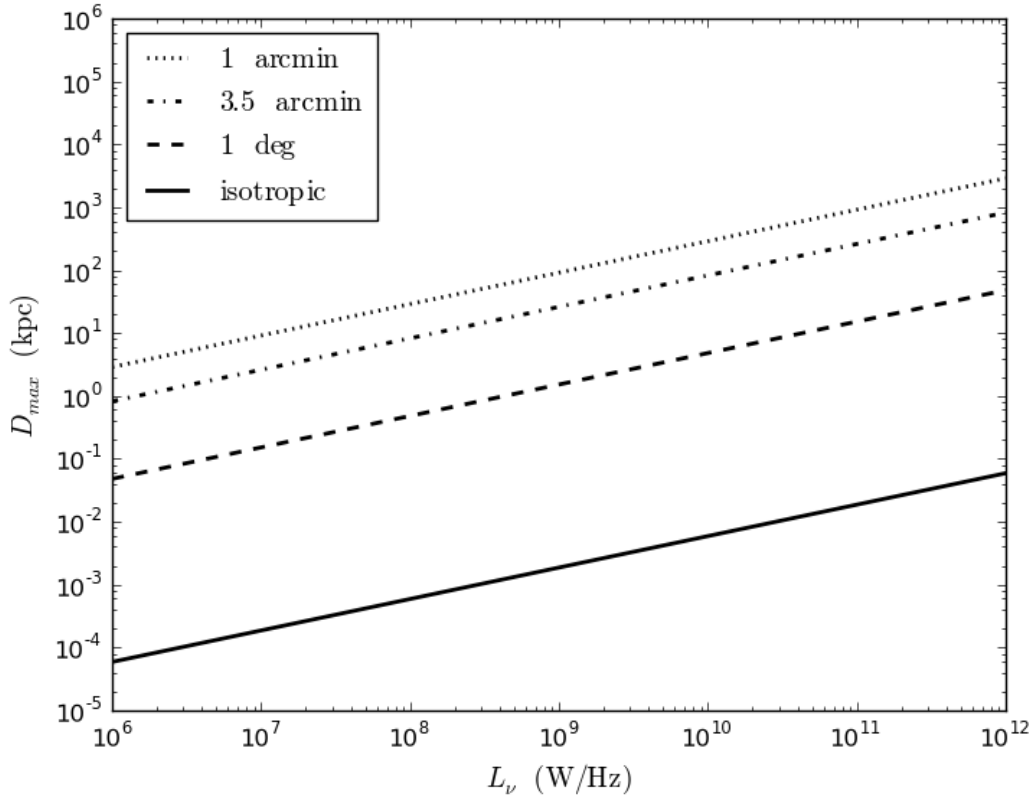


Figure 2.1: The maximum distance that a narrowband signal is detectable with Arecibo using the ALFA receiver and the SERENDIP V.v spectrometer as a function of transmitter power. The bandwidth of the signal is assumed to be 1 Hz. Each curve assumes a minimum detectable flux density of $S_{\min}$ = 30 Jy. The four curves represent different transmitter beaming factors ranging from isotropic (bottom) to a half-power beam-width of 1 arcmin (top) The HPBW of ALFA is 3.5 arcmin.

## 2.3 Survey Description

### 2.3.1 SERENDIP V.v spectrometer

The SERENDIP V.v spectrometer is a digital Fast Fourier Transform (FFT) spectrometer that calculates 134 million ($2^{27}$) channels over 200 MHz of bandwidth, resulting in a spectral resolution of 1.5 Hz. The spectra are not accumulated, so the dump time is 0.67 s. It is comprised of two CASPER[3] FPGA boards: an iBOB and BEE2. An iADC board connected to an iBOB samples ALFA's IF at 800 Msps. The iBOB digitally mixes the signal to baseband and applies a 200 MHz bandpass filter. The signal is then sent to a BEE2 board for a two-stage channelization and realtime thresholding operation. The first channelization stage calculates a coarse spectrum using a 4-tap, 4096 channel polyphase filterbank-FFT. The coarsely channelized data are buffered until 32k coarse spectra (for a total of $2^27$ samples) have been calculated. The buffered data are then transposed and fed into a $2^{15}$ point FFT, which yields the fine, 134 million channel spectra (Siemion et al., 2011).

In order to reduce the data rate, a thresholding operation is performed in real time. Each fine frequency channel is compared to the mean power in its associated coarse frequency bin. If the fine power is higher than the mean power multiplied by a scale factor, the power and bin number of the "hit" is recorded. To further prevent large data recording rates, the number of hits recorded per coarse bin is capped at 25. We also record the mean powers for context. The end data products include the entirety of the coarse spectra, and the sparse spectra comprised of "hits".

[3] https://casper.berkeley.edu/wiki/Main_Page

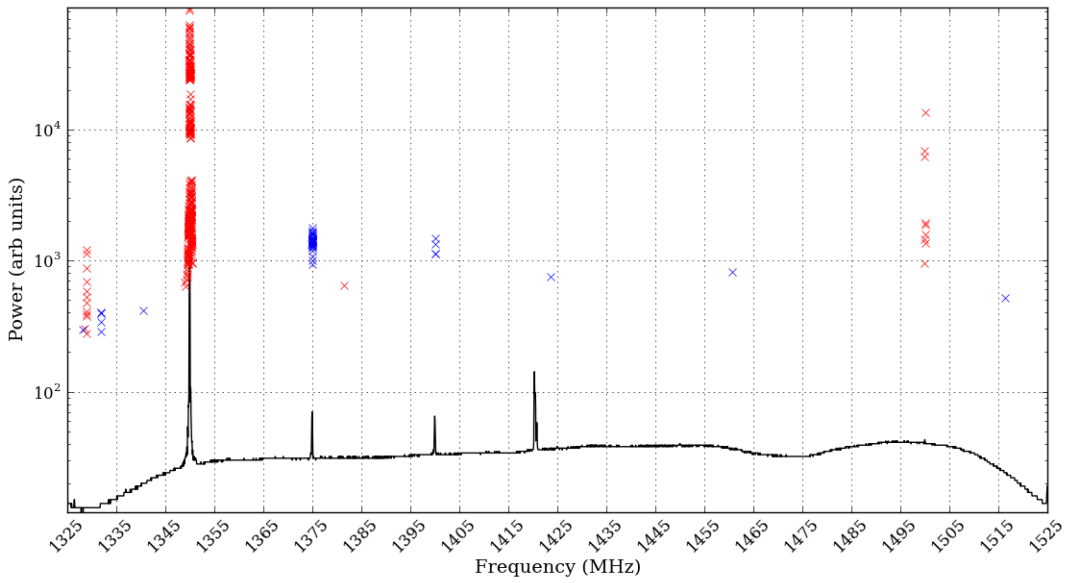


Figure 2.2: Example spectrum from the SERENDIP-V.v spectrometer. The coarse spectrum is shown in black and the fine spectrum hits are shown as crosses. The red crosses are hits occurring at frequencies with known RFI.

Figure 2.2 plots a single spectrum from data taken on 01 May 2010 at a Galactic latitude and longitude of $b = 0.4°$ and $l = 48°$. The 4096-channel coarse spectrum is shown in the solid black line. A spectral line from Galactic hydrogen is clearly visible at a frequency of 1420.4 MHz. The hits are plotted as crosses. The red crosses are hits identified with frequencies of known RFI (see Section 2.3.3). The threshold scale for this spectrum was 20 times the mean power.

### 2.3.2 Observational Setup

The SERENDIP V.v spectrometer records data anytime the Arecibo L-band Feed Array (ALFA)[4] is used for observations. ALFA is a seven-pixel feed array that operates from 1225 - 1525 MHz with dual linear polarization. A single central pixel (Beam 0) is surrounded by a hexagonal ring of six pixels (Beams 1–6). It is employed in several large-scale surveys underway at Arecibo, both galactic and extragalactic, yielding a large sky coverage important for SETI studies. SERENDIP V.v processes the top 200 MHz of the ALFA band. The exact frequency range depends on the observer's choice of RF center frequency. Processing the upper part of the ALFA band avoids RFI caused by the airport radars that plague the bottom of the band.

The spectrometer can only process one of ALFA's 14 RF channels (seven beams, two polarizations) at a time. To regain some of the multi-pixel capability, which is particularly important for RFI rejection, we use an analogy switch to select each of the 14 RF channels in turn. Two spectra are calculated for each RF channel before switching to the next. Both polarizations are observed for a single beam before moving on the the next (i.e. 0A, 0A, 0B, 0B, 1A, 1A, etc. ). The entire cycle completes every 18.8 s.

The SERENDIP V data were taken with several different observing modes according to the scientific needs of the different ALFA surveys. Drift scan surveys fix the telescopes position and map the sky by allowing the sky to pass through the telescope beam. For the 3.5 arcmin ALFA beam, a source crosses the beam in 14 s and the sky drift rate is 0.004 deg/s. Other surveys track a position on the sky, so the sky drift rate is 0.0 deg/s. The data recorder code has no

[4] http://www.naic.edu/alfa/

information about when the telescope is slewing, so the data also contain spectra taken while the telescope is moving. The maximum slew speed corresponds to a maximum sky drift rate of about 0.4 deg/s.

### 2.3.3 Database

The event-based nature of our data lends itself well to the use of a database for data analysis. By constructing the proper database query, a subset of the hits can be selected in frequency, time, beam, etc. Our MySQL database comprises three tables. One table holds information that is valid on a spectrum-by-spectrum basis such as the date and time of the observation, the right ascension and declination of the pointing, the selected RF channel, and the RF center frequency. A "hit" table contains all of the information specific to each event such as the event power, the power of the corresponding coarse channel, the topocentric and barycentric frequencies of the hit, and a RFI flag. The third table contains the entire 4096 channel coarse spectrum. A diagram of the database schema is known in Figure 2.3.

Before inserting a hit into the database, a few data integrity checks are done. If the input RF level is too low, the spectrometer generates a large number of hits. For example if the observer selects a RF center frequency that places the upper part of the SERENDIP V.v band outside of ALFA's band, a large number of hits can be generated in the channels above the bandpass rolloff. To avoid loading these events into the database, we first check that the frequency of each hit is within ALFA's frequency range. If a beam is malfunctioning, a huge number of spurious hits is generated across the entire band. We count the total number of

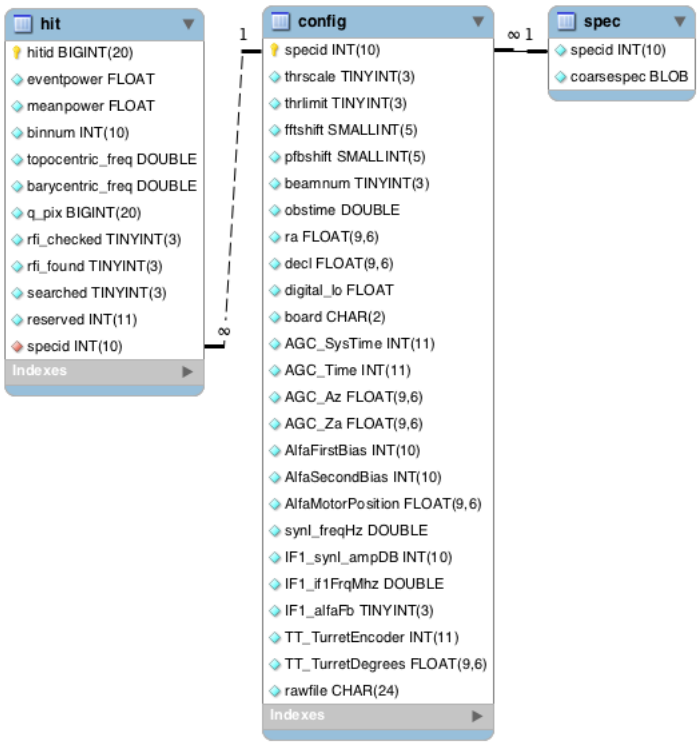


Figure 2.3: Diagram of the SERENDIP V database schema.

hits in a spectrum, and if it is greater than 90% the maximum possible number of hits allowed by our cap, we do not load any of the hits for this spectrum but do insert the information for the spectrum to maintain continuity.

Each spectrum is time tagged with one second resolution by the data recorder code. Because our spectra have a time resolution of 0.671 s, we need to perform an interpolation to estimate a more precise time. This is done by calculating a series of reference times. We can find a spectrum that lines up with the start of a second by noticing that our time resolution, $\Delta t = 0.671$, is 4.4 ms longer than 2/3 seconds. Occasionally the situation arises where one spectrum arrives at the very end of a 1 s time stamp interval, the next two spectra arrive in the latter third and first third of the next two 1 s intervals, and the fourth

spectrum arrives at the beginning of the fourth one-second interval. This sequence of four consecutive one-second time stamps allow us to define reference times. We calculate the Modified Julian date (MJD) for these spectra, and the MJD for all other spectra are calculated by interpolating between the reference times within a continuous block of observing.

Each spectrum is classified by its observing mode, and more specifically by its drift rate ($\dot{\Omega}$). Data taken in different observing modes must be analyzed differently. For each continuous block of observing, the right ascension and declination values are used to calculate a sky drift rate for each spectrum. Four observing modes are defined by ranges of drift rates: $\dot{\Omega}$ <= 0.0001 deg/s are tracked; 0.0001 deg/s < $\dot{\Omega}$ < 0.0039 deg/s are slow drift; 0.0039 deg/s <= $\dot{\Omega}$ <= 0.0043 deg/s are sky drift, and $\dot{\Omega}$ > 0.0043 deg/s are fast drift or slewing. The observing mode for each spectrum is tracked in the database and allows one to easily select the spectra associated with a particular observing mode for analysis.

Through manual inspection, we identified frequencies with high concentrations of hits and compared those frequencies to lists of known narrowband RFI at Arecibo. A list of five frequency ranges that could be correlated with known RFI sources was generated. For example the bright line at 1350 MHz in Figure 2.2 is caused by airport radar. Hits occurring in a frequency range known to contain RFI are flagged using a dedicated column in the database. By constructing the proper query, hits marked as RFI can be ignored in the analysis. This is illustrated in Figure 2.2. Hits plotted in red are associated with known RFI, and only blue crosses are included in further analysis.

## 2.4 Data Description

Commensal observing with ALFA results in data obtained in a variety of observing modes, and these heterogeneous data must be analyzed in different ways. In data taken in drift scan mode, an ETI source crosses the beam (assuming it passes through the center) in about 14 seconds. Since the SERENDIP V.v spectrometer only observes the same beam for four spectra (2.7 s) and requires 19 seconds before it returns again to the same beam, a real signal will only be seen in at most two or four spectra depending on whether the signal is polarized or unpolarized. The successful confirmation of an ETI signal in a drift scan survey would require several re-observations of the same position. As described in Cordes et al. (1997), re-observations separated by longer than the scintillation time scale requires special consideration.

The complementary observing mode tracks a position on the sky. In this case a real signal may be observed multiple times in a single pointing provided the intrinsic variability is longer than the RF channel cycle time. Multiple observations of a signal in a single pointing provides a stronger case than a single observation in drift scan data. We don't expect a large number of re-observations of the same position of the sky in our data, so we have chosen to focus our analysis on data taken in a tracked observing mode.

The majority of the data taken in tracked mode is from the Pulsar ALFA (PALFA) survey (Cordes et al., 2006). PALFA surveys the Galactic plane ($|b| < 5°$) for pulsars in two Galactic longitude ranges: inner galaxy ($30° < l < 78°$) and outer galaxy ($162° < l < 214°$). The duration of a single pointing is typically 268 s, but often half or double that duration. The tracked observing mode in

conjunction with the fact that these observations are made in the Galactic plane with a higher concentration of exoplanets, makes these data particularly interesting.

### 2.4.1 Multi-pixel Mapping

The multi-pixel capability of ALFA provides a powerful RFI rejection tool. A real astrophysical signal may be visible in a single beam or in two or three neighboring beams depending on the relative sky positions of the source and beam centers. On the other hand, RFI is local to the Earth and may be visible in any number of beams simultaneously. Because we can only observe one beam at a time, classifying signals based on multi-pixel information maps to the persistence across several spectra.

Figure 2.4 illustrates three possible RFI signals (red) and an ETI (blue) signal in simulated SERENDIP V.v data. The duration of the simulated data is similar to the duration of two, 134-s pointing in the PALFA survey with a gap of 16 seconds during which the telescope moved to a new sky position. The noise is Poisson distributed in time, uniformly distributed in frequency, and exponentially distributed in amplitude. The dashed vertical lines denote the beginnings for two different pointings. A strong, persistent RFI signal was inserted at a frequency of 1350 MHz. Such a signal can be visible in all beams and therefore all spectra. A second possible RFI signal was inserted at a frequency of 1400 MHz and is an example of variable, narrowband RFI. In this case the signal is visible in Beam 0, 1, and 4, and we can identify this signal as RFI because Beam 1 and Beam 4 are not neighboring. A third class of RFI was inserted at a frequency

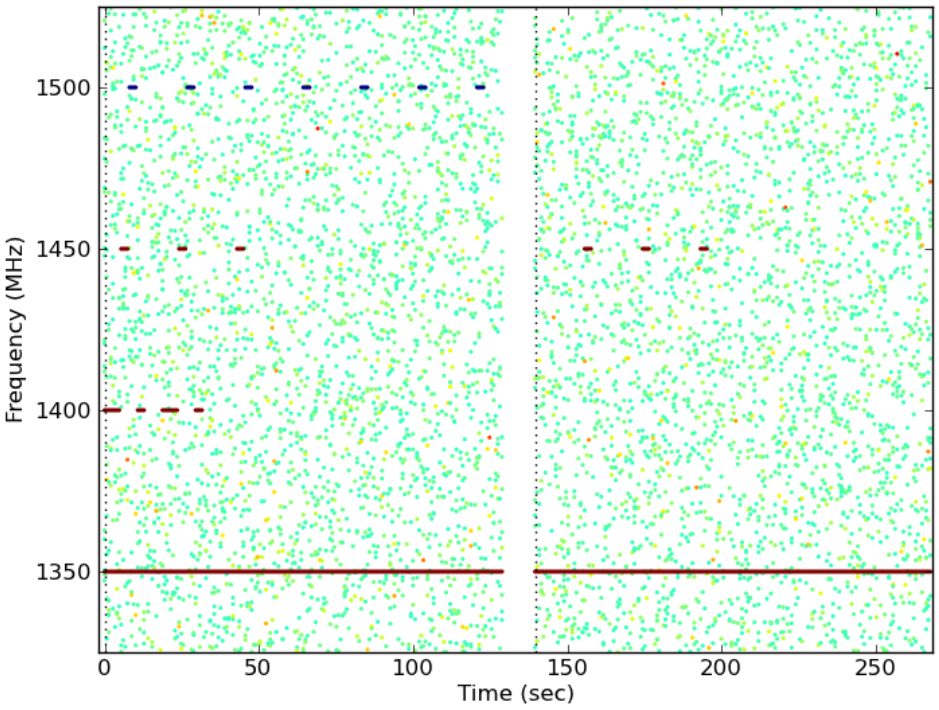


Figure 2.4: Simulated SERENDIP V data. The noise is Poisson-distributed in time, uniformly distributed in frequency, and exponentially distributed in amplitude. Three classes of RFI are shown in red. An ETI signal is shown in blue.

of 1450 MHz. This RFI only appeared in Beam 2, but because it occurred at the same frequency at two different sky positions, we classify it as RFI. Finally a simulated ETI signal was inserted at a frequency of 1500 MHz in Beam 3. The fact that the signal is isolated to a single beam, repeats across many cycles, and is only visible at a single sky position makes it a strong candidate for an astrophysical signal.

## 2.5 Processing Description

### 2.5.1 Finding Pointings

The algorithms described below are applied to the data on a pointing-by-pointing basis and across an entire observing session. A single observing session likely includes several pointings. The boundaries of each pointing are determined using the observing mode parameter described in Section 2.3.3. A drift rate for a tracked survey is 0 deg/s interspersed by short periods with high drift rates while the telescope is moved to a new position. By determining when the drift rate switches from high to low and low to high, we can define the boundaries of each pointing during an observing session. A pointing must be at least 60 s long to be considered valid, and a gap of more then 30 minutes between the end of one pointing and the beginning of the next is assumed to be a new observing session.

### 2.5.2 Clustering Analysis

A time-domain clustering algorithm is used to consolidate hits occurring in consecutive spectra at the same frequency. A cluster comprises a set of hits at a single frequency for which there is a gap in time of at most one spectrum. A cluster is characterized by the frequency, the sample time of the first spectrum in the cluster, the length of cluster in time, and a max or mean power. Hits due to noise will generally have cluster lengths of one sample, so we only consider clusters of lengths longer than one sample in our cross-beam algorithms.

For all the clusters longer than one sample, we look for frequency channels that contain more than one cluster. Because we are focusing on persistent ETI signals, we require that clusters occur at a given frequency at least twice in a single pointing to be considered further. We then determine the unique beam numbers associated with the spectra in clusters and compare them to the allowed number and combinations. If the number of beams is greater than three, the clusters are immediately rejected. If the number of unique beams is two of three, they are compared to a list of allowable combinations and either kept or rejected. A combination of two or three beams is allowed if they are neighboring. For example Beam 0 and any other beam is allowed, but Beam 1 and Beam 5 is not. If the number of beams is a cluster is one, it is kept.

Finally, clusters in individual pointings are compared within the entire observation session. The frequencies of the clusters are grouped into bins of width 100 Hz, and any clusters that occur in more than two pointings in a single frequency bin are rejected as sporadic RFI.

These clustering signal classification algorithms are illustrated in Figure 2.5. The top panel plots the clusters with lengths longer than one sample for the simulated data shown in Figure 2.4. The vertical tick denotes the center of the cluster and the horizontal bars mark the time span. The middle panel illustrates the results after applying the cross-beam analysis. The persistent RFI at 1350 MHz was rejected because it comprised spectra that came from more than three beams. The variable RFI at 1400 MHz occurred in three beams, two of which were not neighboring and also rejected. The sporadic RFI that only occurred in a single beam and the ETI signal both pass these tests. The bottom panel illustrates the results after also applying the multi-position analysis. For illus-

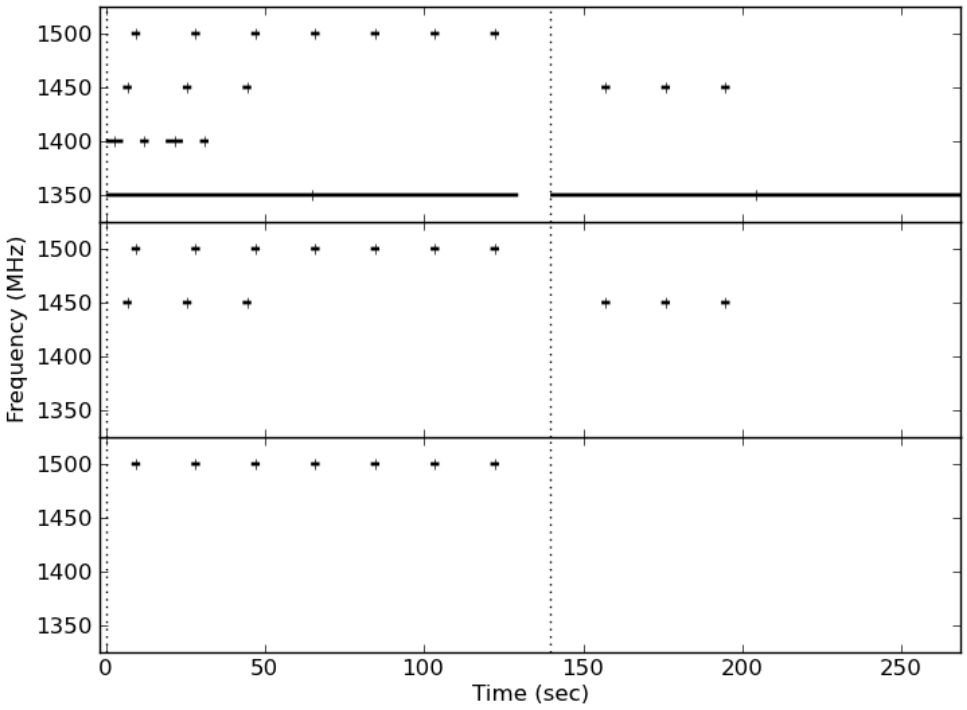


Figure 2.5: Cluster analysis of fake SERENDIP V shown in Figure 2.4. The top panel shows all clusters with lengths longer than one sample. The middle panel shows the clusters remaining after the cross-beam analysis. The bottom panel shows that only clusters from the ETI signal survive after applying the multi-position rejection algorithm.

trative purposes, we only allow a candidate ETI signal to occur in a single beam, where as in our analysis of the real data, we allow a signal to occur in up to two pointings. The remaining RFI is rejected because it is present in both pointings.

### 2.5.3 Reobservations

Once a list of candidate clusters is generated, we search for candidate ETI signals by comparing clusters seen in the same region of the sky, at nearby frequencies, and in different observing sessions. The angular offset between the beam center of a reference cluster and the beam center of every other cluster is calculated, and offsets less than seven arcmin are identified. An offset of seven arcmin allows for the possibility of detections in the first sidelobe. Furthermore our cross-beam analysis allowed for a signal to be visible in up to three beams as long as they were neighboring, and an offset of seven arcmin includes the regions where this is geometrically possible.

If clusters spatially close to the reference cluster are found, the difference in observation time (MJD) and barycentric frequency between each of these clusters and the reference cluster is calculated. Clusters with a difference in MJD of greater than 30 min and frequency separations of less than 150 kHz are recorded. The time difference of 30 min was chosen because 30 min the minimum time allowed between pointings before a new observing session is declared (see Section 2.5.1). The maximum Doppler shift for Earth orbiting the Sun is 140 kHz (see Section 2.2.1), so we allowed a given frequency to have drifted by at most 150 kHz to be considered to have possibly come from an exoplanet.

With the group of clusters all observed in the same sky position, we look at how the cluster barycentric frequency varies in time. If a signal is being transmitted from an ETI, the barycentric frequency should vary sinusoidally with a period commensurate with the planet's orbital period. Such a determination requires several reobservations of the same sky position.

## 2.6 Results

In this work we present the results from the data collected in 2010. Manually we identified raw data files that contain data from PALFA, as well as any observations for which we did not know the observing mode, and inserted these into our database. A total of 812,633 spectra were recorded during targeted observations for total time on sky of 6.3 days. The number of hits in these spectra was 334,320,825.

The clustering analysis described in Section 2.5.2 was applied to the data. First a set of pointing boundaries were determined as described in Section 2.5.1. The valid hits within the range of spectra within a single pointing are chosen. A hit is valid if it has not been flagged as RFI, is associated with targeted observing, and has a mean power greater than 150. Roughly half of these hits are flagged as being associated with frequencies of known RFI. The last condition is included because low mean power is associated with a large number of spurious hits. The value of 150 was chosen empirically. About 11% of the hits remain once all three conditions are applied.

The valid hits in each pointing are converted to clusters and the cross-beam RFI identification algorithm is applied. Around 457,000 clusters were identified in the roughly 39,000,000 valid hits. For all of the pointings associated with a single observing session, the multi-pointing RFI rejection algorithm is applied. The number of clusters remaining after the multi-pointing algorithm is about 41,300. The properties of the cluster candidates determined for a single pointing and the clusters that pass the multi-pointing algorithm are written to text files, and a dynamic spectrum (frequency vs. time) diagnostic plot is generated. An

example of such a plot is shown in Figure 2.6.

Figure 2.6 shows a dynamic spectrum of the clustering algorithms applied to two hours of PALFA observations from 1 May 2010. The topocentric frequency of each cluster is plotted versus the time of the cluster relative to the start of the plot. The clusters that passed through the cross-beam algorithm that acts on a pointing-by-pointing basis are plotted as black dots and shown for context. The clusters that also passed through the multi-pointing algorithm are shown color coded by the beam number of the first sample in the cluster. Red is Beam 0, and blue is Beam 6. The vertical dotted lines delineate the beginning of a new pointing. For this particular observing session, the duration of a single pointing was about 530 s. Despite both the cross-beam analysis and cross-pointing analysis, a large amount of RFI is still visible at frequencies less than 1350 MHz. There is a line of clusters near 1375 MHz that were not fully removed. Most other clusters were found at isolated frequencies.

Figure 2.7 shows the positions of all of the clusters in our analysis. The dashed black line outlines the Galactic plane ( $|b| < 5^\circ$ ). The concentration of observations in the Galactic plane with an further emphasis on low Galactic longitudes ($18 < RA < 21$), follows from our focus on data collected during PALFA observations.

Finally we group clusters by reobservation as described in Section 2.5.3. For each cluster in our list, we find all others that are spatially nearby (angular offset of less than seven arcmin), nearby in frequency (frequency difference of less than 150 kHz), and separated in time by more than 30 min. This approach leads to many groupings that are similar, as pairs will be found more than once as each one used as the reference cluster in turn. When the grouping algorithm

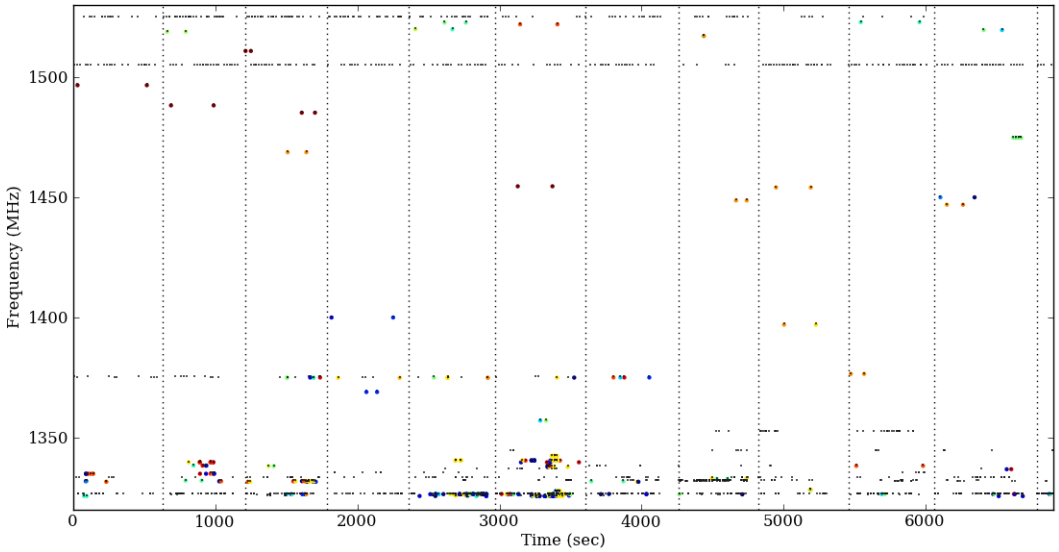


Figure 2.6: Dynamic spectrum of clusters from a two hour session of PALFA observing. The topocentric frequency of the cluster is plotted versus the time. All of the clusters that passed the cross-beam RFI rejection algorithm are shown as black dots. The clusters that also passed the multi-pointing rejection algorithm are shown as colored circles. The color of the circle represents the beam of the first sample in the cluster. The vertical, dotted lines show the beginning of each pointing.

was applied to our cluster list, 95% of the groupings occurred at frequencies below 1350 MHz, which as was shown in Figure 2.6, is still riddled with RFI. Therefore we only consider groupings at frequencies above 1350 MHz hereafter.

Most of the cluster groupings include clusters observed in only two independent observing sessions. Two measurements is not constraining, so we focus on groupings that include three or more clusters from different observing sessions. Figure 2.8 illustrates two cluster groups with members from more than two independent observing sessions. A barycentric frequency offset is plotted versus

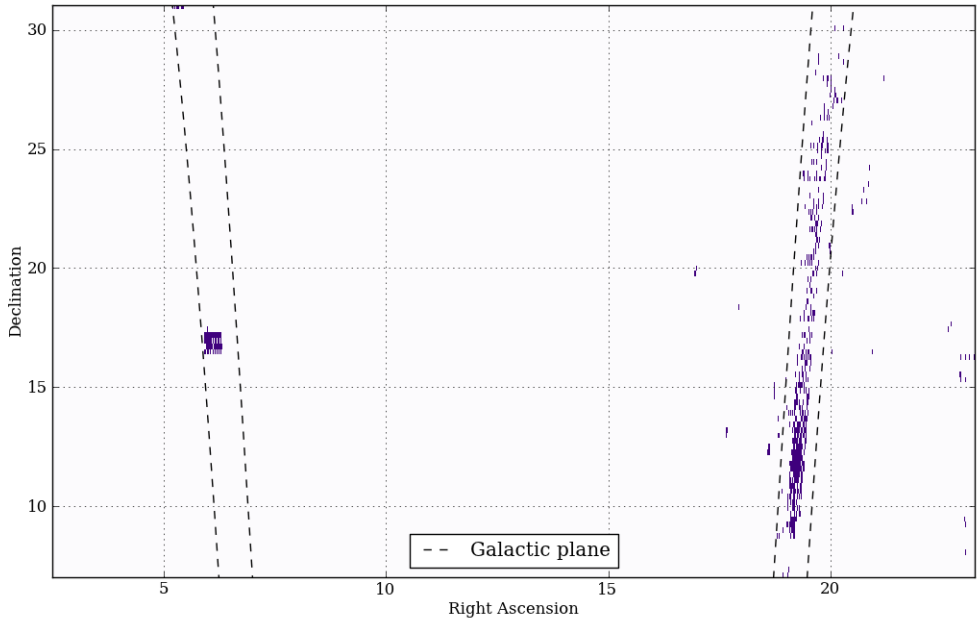


Figure 2.7: Right ascension and declination of the clusters included in our analysis. The dashed black line denote the Galactic plane ($|b| < 5°$). Most of our data correspond to observations in the Galactic plane with a greater emphasis on lower Galatic longitudes (18 h < RA < 21). This is expected as we focused on data collected during the PALFA survey.

modified Julian date (MJD). An offset is plotted for clarity and the value of the frequency offset is given in the upper left corner of each panel. The cluster group in the top panel shows an overall increase in frequency of about 5 kHz over about 100 days. The cluster group in the bottom panel shows a general decrease in barycentric frequency over about 120 days. In both cases it would be difficult to constrain possible orbital models with so few points.

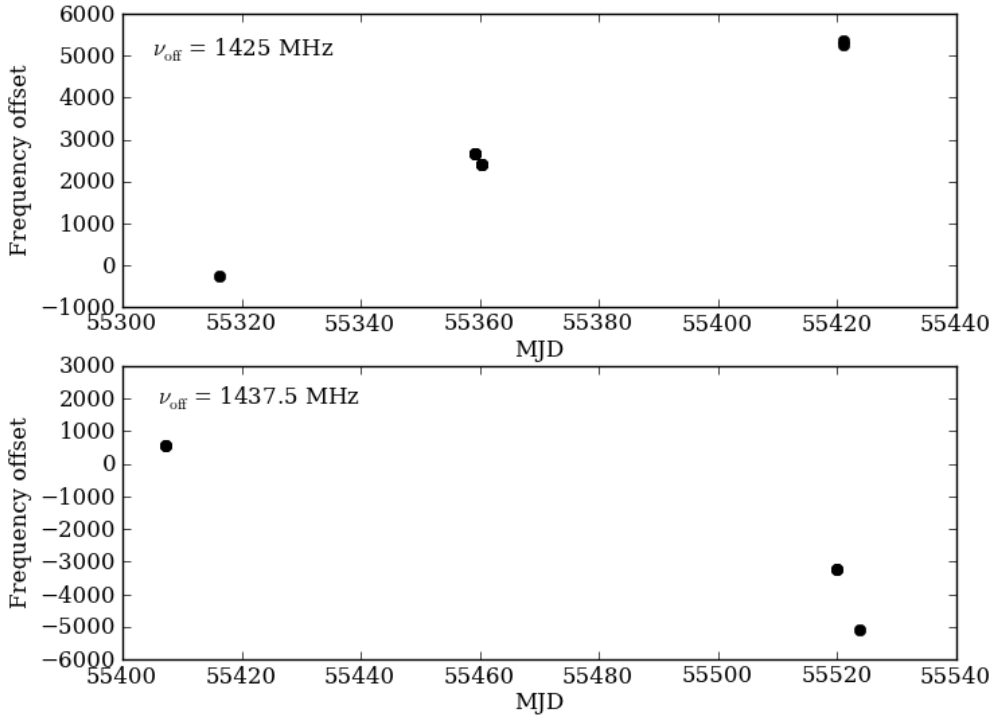


Figure 2.8: Barycentric frequency vs. modified julian date (MJD) of groups of clusters spanning more than three observing sessions. For clarity we've plotted the barycentric frequency with an offset subtracted. The offset for each case is given in the upper left corner of the panel.

## 2.7 Discussion

Radio SETI experiments have to balance acquiring as much information and data as possible and only acquiring as much as can be digested. For example SETI@home records the raw voltages directly, which allows one to explore the data to its full potential, but is limited to searching a narrow bandwidth out of practical considerations (hardware, CPU time, storage capacity). The SERENDIP experiments observe over a wider bandwidth and mitigate the prac-

tical challenges by limiting what information is kept through a real-time thresholding operation. As such, one is limited in what one can gleam from the data alone and must use additional information. In particular the biggest difficulty is distinguishing a hit as arising from noise, RFI, or a real, ETI signal. In this analysis we include auxiliary information such as a list of frequencies known to be regularly contaminated with RFI and cross-beam anti-coincident RFI rejection.

Our experience from these data inform two suggestions that any future SERENDIP-style experiment should follow:

**Simultaneous multi-pixel observations** Efficiently identifying and excising RFI is crucial, and a multi-pixel system would enable simultaneous cross-beam RFI rejection. For example, if we had a fourteen-pixel spectrometer on ALFA, we could identify RFI that varies on time scales shorter than our RF cycle time. Arecibo's optics is complicated, and often RFI varies in amplitude across the seven beams. Simultaneous measurements of all seven beams would more efficiently identify these case than our time-based cross beam analysis. Furthermore having a multi-pixel spectrometer on ALFA would also increase the survey speed by a factor of seven.

**Many reobservations of the same position** The positive identification of an genuine ETI signal will require many reobservations. If the ETI signal is a beacon being transmitted at a fixed frequency then the observed frequency should shift as the exoplanet (or exomoon) orbits its star (or planet). But the initial identification of candidates to follow up more regularly also requires many observations of the same sky position. In our experiment commensal observing on ALFA assures us a lot of observation time covering a large fraction of the

sky visible from Arecibo, but we do not have a lot of reobservations of the same position, as most large surveys only cover a given position a couple of time. Ideally a SERENDIP-style project would operate on a telescope with a large survey speed (i.e. large instantaneous field of view) and in an observing mode that allows frequency reobservations. Here again a multi-pixel spectrometer system would be valuable.

## 2.8 Conclusions

We have presented results from the SERENDIP V survey operating at the Arecibo Observatory. This SETI experiment operates commensally on the seven-pixel ALFA receiver system. Data are recorded with the SERENDIP V.v spectrometer, a single-pixel FFT spectrometer, with a very high frequency resolution. Our analysis of the data relies on a time-domain clustering algorithm to consolidate hits that have the same frequency and are adjacent in time into clusters. A cross-beam RFI rejection is applied to the clusters to remove those that comprise non-neighboring beams. The clusters that pass the cross-beam rejection algorithm are then also subjected to an algorithm that checks for clusters found at similar frequencies in different sky positions as these are likely RFI. The clusters that remain are grouped by sky position, barycentric frequency, and observing time. No ETI signal was identified, but most sky positions were observed only once and only a handful more than twice. Future SERENDIP-style experiments should employ a multi-pixel receiver and spectrometer system and an observing strategy with many passes of the same sky positions.

CHAPTER 3

# MULTIMOMENT RADIO TRANSIENT DETECTION

We present a multimoment technique for signal classification and apply it to the detection of fast radio transients in incoherently dedispersed data. Specifically, we define a spectral modulation index in terms of the fractional variation in intensity across a spectrum. A signal whose intensity is distributed evenly across the entire band has a lower modulation index than a spectrum whose intensity is localized in a single channel. We are interested in broadband pulses and use the modulation index to excise narrowband radio frequency interference (RFI) by applying a modulation index threshold above which candidate events are removed. The technique is tested both with simulations and using data from known sources of radio pulses (RRAT J1928+15 and giant pulses from the Crab pulsar). The method is generalized to coherent dedispersion, image cubes, and astrophysical narrowband signals that are steady in time. We suggest that the modulation index, along with other statistics using higher-order moments, should be incorporated into signal detection pipelines to characterize and classify signals.[1]

## 3.1 Introduction

Surveys have always played an important role in astronomy and will play an increasingly important role in the future as observatories such as the Large Synoptic Sky Telescope (LSST) and the Square Kilometer Array (SKA) come online. The huge volumes of data generated by surveys require robust pipelines to identify and characterize populations with maximal completeness and mini-

[1]Published in the *Astrophysical Journal* as Spitler et al. (2012).

mal false positives. Regardless of the target source class, all detection pipelines rely on signal-to-noise ratio (SNR) to find sources and quantify their believability. This approach underutilizes the spectral information contained in the data, because it only uses the total intensity, or first moment, of a spectrum. We propose adding a second statistic, the spectral modulation index, that uses both the first moment (signal mean) and second moment (signal variance) of a spectrum to classify signals found through their SNR.

Although the technique we present in this paper is applicable to any data collected as intensity versus time and frequency, we focus on surveys for fast radio bursts and use the modulation index to identify and remove narrowband radio frequency interference (RFI). Surveys at radio frequencies must contend with RFI because, if not mitigated or excised, it can produce many false positives in processing pipelines. One difficulty in removing RFI is that it arises from a wide variety of terrestrial sources with different signal characteristics. Moreover it can be episodic or simply transient in nature along with the astrophysical signals that we are interested in. RFI can be broad in time and narrow in frequency (e.g. Global Positioning System satellites) or broad in frequency and narrow in time (e.g. lightning). Some radar signals sweep in frequency and mimic the plasma dispersion of astrophysical bursts.

We define fast radio bursts as having characteristic widths less than about one second, so astrophysical plasma delays, such as those encountered in pulsar signals and in the class of transients known as rotating radio transients (RRATS, McLaughlin et al., 2006), are important. When applied to fast transients, our technique builds upon the methods presented in Cordes & McLaughlin (2003) and McLaughlin & Cordes (2003) and that led to the discovery of RRATs. How-

ever, the basic idea applies to transients of any duration.

Actual signals will blur the distinction between RFI and signals of interest because some RFI will be broadband and some broadband astrophysical signals will show significant frequency variation. Astrophysical sources may have a spectral dependence including a simple spectral index or stronger modulations like those seen in solar bursts. Compact sources of fast transients will typically show frequency modulation from interstellar scintillation. We consider these effects in our implementation of the method.

The modulation index is not a new statistic; it has been used to measure time variations in a variety of astronomical applications, such as variability of pulsars (Weisberg et al., 1986) and extragalactic sources (Kedziora-Chudczer et al., 2001) caused by interstellar scintillations. It has also been used to study properties of the solar wind (Spangler & Spitler, 2004).

We lay out the mathematical groundwork for multimoment dedispersion and the calculation of the modulation index in § 3.2. In § 3.3 we discuss the implementation of the modulation index in a fast transient detection pipeline and present the results of a simulated single pulse detection pipeline. Also in § 3.3 we apply the technique to two known sources of single pulses: Crab giant pulses and RRAT J1928+15. The method is extended to other types of data sets in § 3.4. We discuss how the spectral modulation index can be used to classify bursts from a variety of astrophysical sources in § 3.5 and make concluding remarks in § 3.6.

## 3.2 Method

For specificity we consider broadband astrophysical signals that are sampled as dynamic spectra; that is, a sequence of spectra separated in time by $\Delta t_s$ with $N_\nu$ frequency channels spanning a total bandwidth $B$. Most of the cases we discuss in this paper will have time-bandwidth products well in excess of unity, i.e., $\Delta t_s B/N_\nu \gg 1$, as is consistent with fast-dump spectrometers used in surveys for pulsars and radio transients. Our discussion will also concentrate on incoherent (post-detection) dedispersion, although we briefly discuss applications where coherent dedispersion is used. To illustrate the basic method, we consider only a simple sum over frequency to yield an intensity time series; later we will consider interstellar dispersion delays in the sum and usage of the dispersion effect in discriminating astrophysical signals from RFI.

We define the modulation index $m_I$ as the normalized standard deviation of the intensity $I$ across frequency ($\nu$),

$$m_I^2 = \frac{\overline{I^2} - \overline{I}^2}{\overline{I}^2}, \tag{3.1}$$

where the first and second moments, $\overline{I}$ and $\overline{I^2}$, respectively, are given by

$$\overline{I^n} = N_\nu^{-1} \sum_\nu \mathrm{I}^\mathrm{n}(\nu). \tag{3.2}$$

The modulation index, $m_I$, characterizes the distribution of signal power in spectrum. If the power is distributed evenly across the band, the variance is small, and the spectrum has a low modulation index. A broadband pulse with a flat spectrum and infinite signal-to-noise ratio (SNR) has $m_I = 0$. In the opposite extreme where the power is localized to single spectral channel, the modulation index is $m_I \rightarrow \sqrt{N_\nu - 1}$ with increasing SNR. These simple examples illustrate

how broadband astrophysical transients can be discriminated from RFI, which often consists of narrow spikes in frequency accompanied by time variability, by requiring that the modulation index be less than some ceiling, $m_{\mathrm{I,max}}$.

### 3.2.1 Multimoment Dedispersion

The modulation index improves upon current detection schemes by including more information about the signal through the calculation of both the first and second moments. The signal processing algorithms used in the detection of short-duration radio transients (i.e., dedispersion and smoothing) must therefore be expanded to higher order moments. Although we only use the first and second moments in this paper, one could consider statistics that use higher order moments (e.g., skewness and kurtosis), so we introduce general, $n^{th}$-order expressions.

Radio pulses traveling through the interstellar medium are subject to frequency-dependent dispersion. Standard pulsar processing techniques remove the effects of dispersion by shifting the intensity in each frequency channel in time according to the $\nu^{-2}$ dispersion relation and averaging to increase the signal-to-noise ratio (SNR).

The standard incoherent or post-detection approach for calculating a dedispersed time series is

$$\bar{I}(t, \mathrm{DM}) = \frac{1}{\mathrm{N}_\nu} \sum_{\nu} \mathrm{I}(\mathrm{t} + \mathrm{t}_{\mathrm{DM}}(\nu), \nu) \tag{3.3}$$

where $I(t, \nu)$ is the time-frequency intensity data, $N_\nu$ is the number of frequency channels, and $t_{DM}(\nu)$ is the delay at frequency $\nu$ for dispersion measure DM.

Throughout we will denote an average in frequency with a bar over the variable. Survey data must be dedispersed using a set of trial DMs because the DM of the target sources are not known *a priori*, except in special applications, such as searches of globular clusters with previously known pulsars. A list of candidate pulses is defined by applying a minimum SNR threshold ($\mathrm{SNR_{min}}$) to the set of dedispersed time series.

Equation 3.3 implicitly weights all frequency channels equally. In general though there will be effects, both astrophysical and instrumental, for which optimal detection (i.e., maximum SNR) requires unequal weighting. For example broadband astrophysical sources have spectra with slopes characterized by a spectral index $\alpha$, and maximum SNR occurs when the frequency channels are given some weighting $w_\nu$ that reflects $\alpha$. Similarly removing instrumental effects, like bandpass subtraction, may introduce channel-dependent root-mean-square noise. The generalized expression for a weighted, dedispersed, first moment time series is

$$\bar{I}(t, \mathrm{DM}; \alpha) = \frac{\sum_\nu \mathrm{w}_\nu \, (\nu/\nu_0)^\alpha \, \mathrm{I(t + t_{DM}}(\nu), \nu)}{\sum_\nu \mathrm{w}_\nu}. \tag{3.4}$$

A full discussion of frequency weights is presented in Section 3.2.2.

Optimization in detection can also be made by considering the duration of the pulse relative to the time resolution of the data. The effective time resolution can be decreased by smoothing the data, and the maximum SNR occurs when the effective time resolution of the smoothed data matches that of the pulse (Cordes & McLaughlin, 2003). The simplest method of smoothing adds adjacent time samples until all of the signal's power is in a single sample. In general any smoothing technique can be defined by applying smoothing weights $w_t$ to the

time-frequency data and averaging in time;

$$I_s(t,\nu) = \frac{\sum_{t'} w_{tt'} \mathrm{I(t',}\,\nu)}{\sum_{t'} w_{tt'}}, \tag{3.5}$$

where $I_s(t,\nu)$ is the smoothed time-frequency data and $I(t',\nu)$ is the original time-frequency data. A dedispersed, smoothed time series $\overline{I_s}(t,DM)$ is calculated according to Equation 3.3 substituting $I_s$ for $I$.

The calculation of the modulation index requires the second moment of the intensity time series. We make one final generalization by expanding Equations 3.4 and 3.5 to higher order moments. The weighted, dedispersed $n^{th}$-moment time series is

$$\overline{I^n}(t,\mathrm{DM};\alpha) = \frac{\sum_{\nu} \mathrm{w}_\nu \left[(\nu/\nu_0)^\alpha \, \mathrm{I(t + t_{DM}}(\nu),\nu)\right]^\mathrm{n}}{\sum_{\nu} \mathrm{w}_\nu}. \tag{3.6}$$

The smoothed, dedispersed $n^{th}$-moment time series is

$$\overline{I_s^n}(t,DM) = \frac{1}{N_\nu} \sum_{\nu} \left[ \frac{\sum_{t'} w_{tt'} \mathrm{I(t' + t_{DM}}(\nu),\nu)}{\sum_{t'} w_{tt'}} \right]^\mathrm{n}. \tag{3.7}$$

Because the spectral modulation index is a measure of the frequency structure, and not the temporal structure, the data are smoothed first in time, thereby consolidating the signal into a single spectrum, before it is squared and averaged. The modulation index for smoothed data is then

$$m_I^2 = \frac{\overline{I_s^2} - \overline{I_s}^2}{\overline{I_s}^2}. \tag{3.8}$$

We have specified separate expressions for the dedispersed intensity with frequency weights and smoothing weights solely for clarity; an optimal detection scheme would employ both. In the remainder of the paper we will explore

the role of smoothing on the modulation index but keep the assumptions that $w_\nu = 1$ and $\alpha = 0$.

An alternative formulation to calculating the smoothed, dedispersed time series is to square the data before smoothing:

$$\overline{I_r^n}(t, DM) = \frac{1}{N_\nu} \sum_\nu \frac{\sum_{t'} w_{tt'} \mathrm{I^n(t' + t_{DM}(\nu), \nu)}}{\sum_{t'} w_{tt'}}. \tag{3.9}$$

This approach captures both the time and frequency variation of the data encompassed by the sumations. Throughout this paper we will focus on the spectral modulation index but will discuss this "time resolved" modulation index in Section 3.2.6 when we discuss the signature of incorrect dedispersion.

### 3.2.2 Frequency weights

As discussed in the previous section, weighting the frequency channels non-uniformly when calculating a dedispersed time series can improve the SNR of the detection. In general an astrophysical signal will have a power law spectrum

$$P_{\nu t} = P_{\nu_o t} \left( \frac{\nu}{\nu_o} \right)^{-\alpha_o} \tag{3.10}$$

where $P_{\nu t}$ is the pulse flux density at time $t$ and channel $\nu$, $\nu_o$ is a reference frequency, and $\alpha_o$ is the source's spectral index. Frequency weights that yield the best SNR act to flatten the spectrum

$$w_{\nu,\alpha} = \left( \frac{\nu}{\nu_o} \right)^{\alpha} \tag{3.11}$$

where optimal detection occurs for $\alpha = \alpha_o$. Uniform spectral index frequency weights implicitly assumes a flat spectral index of $\alpha = 0$. Pulsars have typical

spectral indices of $\alpha_o = 1.6$ with significant variation ($\alpha_{o,min} = 0$ and $\alpha_{o,max} = 3$) (Lorimer et al., 1995). Processing with an implied spectral index of $\alpha = 0$ is only ideal for the minority of pulsars with the lowest observed spectral indices. Instead, applying frequency weights corresponding to the mean pulsar spectral index would increase the SNR with little additional computational cost.

Surveys for new classes of sources where $\alpha$ is unknown or surveys searching for extremely weak examples of a known population may warrant a search over spectral index. Dedispersing data using a set of trial spectral indices would increase the required computation by a factor equal to the number of trial spectral indices. Such a search may only be practical for offline post-processing.

Frequency weights can also reflect frequency-dependent noise variations caused by instrumental effects. In general we can define a signal model

$$I_{\nu t} = b_\nu (T_{\nu t} + g_{\nu t} P_{\nu t}),$$

where $b_\nu$ is the bandpass shape, $T_{\nu t}$ is the system temperature, $g_{\nu t}$ is the gain (e.g., K Jy$^{-1}$) and $P_{\nu t}$ is the pulsar or transient flux density. The quantity we are interested in is $P_{\nu t}$, but the quantity we measure is $I_{\nu t}$. Isolating $P_\nu$ requires removing the three frequency-dependent instrumental effects, which may introduce frequency-dependent rms noise. For example flattening the bandpass ($b_\nu$) will result in higher noise at the band edges. For systems operating at low frequencies ($\sim$ 100 MHz) and large total bandwidths, the system temperature may vary across the band due to the strong frequency dependence of the sky brightness temperature. Finally gain variations will be both time and frequency dependent due to the particularities of the instrument. Again one chooses $w_\nu$ such that SNR is maximized, so if the additive noise can be modeled as Gaussian white noise, weighting the channels by the inverse of their variance (i.e.,

$w_\nu \propto 1/\sigma_\nu^2$) results in the maximum SNR.

To estimate the importance of considering frequency-dependent noise variations, we adopted a simple model for system temperature $T_{\nu t}$ = 20K + 10K$(\nu/\nu_o)^{-2.7}$ and a signal with a non-zero spectral spectral index. The resulting SNR after correcting for the spectral index and averaging over frequency was compared to the standard case that assumes a flat spectral index. At higher observing frequencies ($\sim$ 1 GHz) the improvement is small ($\sim$ 0.1%) for small or large bandwidths because of the low sky brightness temperature. At lower observing frequencies ($\sim$ 100 MHz), there is an improvement in SNR by as much as $\sim$ 80% for wide bandwidths. This is particularly relevant for observatories like the Low Frequency Array (LOFAR, de Vos et al., 2009) and Murchison Widefield Array (MWA, Lonsdale et al., 2009).

A non-zero spectral index will increase the variance of a spectrum and therefore also increase the modulation index. This is undesirable because it would mistakenly imply a lower filling factor and may result in a true signal being flagged as RFI. The magnitude of the increase depends on the observation frequency ($\nu_o$), bandwidth ($B$), and spectral index. For $B/\nu_o \sim 0.1$ the modulation index increase is of order $\sim$ 0.1 for $\alpha \sim 3$. In the extreme case of $B/\nu_o \sim 1$, the modulation increase is of order $\sim$ 1 for $\alpha \sim 3$. Current instruments have $B/\nu_o \sim 0.1$, so the increase in modulation index due to a source's spectral index is negligible. The trend is for new instruments to have larger bandwidths, so eventually the modulation increase will be significant enough that correcting for the spectral index becomes necessary. Furthermore in surveys that employ a spectral index search, the modulation index aids the analysis, because the trial spectral index closest to the true spectral index has the lowest modulation index.

### 3.2.3 Modulation Index

The modulation index is a quantitative measure of the patchiness, or modulation, of a spectrum. It differentiates signals whose power is distributed evenly across the band from those whose power is isolated to a few frequency channels. Broadband and narrowband signals have small and large modulation indices respectively. The level of modulation is parametrized by a frequency filling factor $f_\nu = W_\nu/N_\nu$ where $W_\nu$ is the number of channels in the spectrum that contains signal. Broadband astrophysical signals have a high filling factor ($f_\nu = 1$). RFI can be both narrowband or broadband with a filling factor ranging from $f_\nu = 1/N_\nu$ to 1. As this paper focuses on detecting broadband signals, the ultimate goal is to define a modulation index cutoff, $m_{\mathrm{I,max}}$, that will enable us to flag signals with low filling factors and vastly reduce the number of candidates created by a signal detection pipeline.

The analysis below assumes that the data have been searched for candidate signals by requiring a candidate to have an $\mathrm{SNR_t}$ larger than a minimum SNR ($\mathrm{SNR_{min}}$), and the modulation index is only calculated for these candidate signals. Because we assume that our data has zero mean, either through construction in the case of simulated data or through bandpass subtraction in the case of real data, this assumption assures that $\bar{I} > 0$ and $m_\mathrm{I}$ is well-defined. Furthermore, our interpretation of $m_\mathrm{I}$ assumes $\mathrm{SNR_{min}} > 1$, which is reasonable assumption since such a low $\mathrm{SNR_{min}}$ would result in a deluge of events.

To derive a simple, analytical expression for the dependence of $m_\mathrm{I}$ on $f_\nu$, we consider a spectrum with $N_\nu$ channels that contains two components: noise and signal. The noise is assumed to have a mean of zero and variance $\sigma_x^2$. The signal fills $W_\nu$ channels with intensity $A_i$ in channel $i$. While each $A_i$ may be

different, it will be useful to define $\overline{A}$, the average signal intensity over $W_\nu$ channels. The average signal intensity over the entire band is $f_\nu \overline{A}$. As we are usually more interested in signal-to-noise ratios (SNR), we define a single-channel SNR, $\mathrm{SNR}_{\nu t} = \overline{\mathrm{A}}/\sigma_\mathrm{x}$, and time series SNR, $\mathrm{SNR}_\mathrm{t} = \sqrt{N_\nu} \mathrm{f}_\nu \mathrm{SNR}_{\nu t}$. The latter expression assumes that the standard deviation of the noise in the times series is $\sigma_x / \sqrt{N_\nu}$.

The modulation index for this simple signal model is

$$m_\mathrm{I}^2 = \frac{N_\nu}{\mathrm{SNR}_\mathrm{t}^2} + \frac{m_\mathrm{A}^2}{f_\nu} + \frac{1 - f_\nu}{f_\nu} \tag{3.12}$$

where we have introduced a separate signal modulation index, $m_\mathrm{A} = \sigma_\mathrm{A}/\overline{A}$. This modulation index characterizes the inherent frequency structure of a signal, and for this initial discussion we consider only signals with negligible structure ($m_\mathrm{A} \ll 1$). In Sections 3.2.4 and 3.2.5 we discuss the effects of interstellar scintillation and pulsar self-noise and will introduce a non-zero $m_A$.

To explore the behavior of the modulation index as a function of $\mathrm{SNR}_\mathrm{t}$ and $f_\nu$, we will look at the extrema of $f_\nu$: $f_\nu = 1$ and $f_\nu = 1/N_\nu$. It is important to note that Equation 3.12, as well as the limiting expressions defined below, represents the ensemble average values of the modulation index. Statistically they are the average value expected for a given combinations of $\mathrm{SNR}_\mathrm{t}$ and $f_\nu$, but noise in the data will cause variation in the actual values. The expressions below are also idealized in so far as we have set $m_\mathrm{A} = 0$.

We also simulated a time-frequency data set containing broadband and ultra-narrowband signals added to Gaussian noise to accompany the discussion. It was processed according to standard pulsar processing techniques, and the results are shown in Figure 3.1. The data set has $N_\nu = 256$ and includes 100 broadband pulses ($f_\nu = 1$) with $W_\mathrm{t} = 1$ and $\mathrm{SNR}_\mathrm{t} = 10$, two sets of 100 ultra-narrowband spikes ($f_\nu = 1/N_\nu$) with $W_\mathrm{t} = 1$ and $\mathrm{SNR}_\mathrm{t} = 5$ and $\mathrm{SNR}_\mathrm{t} = 10$

respectively, and $10^6$ noise-only spectra with zero mean and $\sigma_x^2 = 1$. We ignored dispersion for simplicity, and the time series was calculated using Equation 3.3 with $t_{\mathrm{DM}}(\nu) = 0$. An intensity threshold was applied to the resulting time series with $\mathrm{SNR}_{\mathrm{min}} = 3$ (solid horizontal line), and the modulation index was calculated for samples above threshold. The vertical dashed line and dashed curve are explained below.

A broadband signal ($f_\nu = 1$) with no inherent structure does not increase a spectrum's variance, so the modulation index depends only on the number of frequency channels and the signal's $\mathrm{SNR}_{\mathrm{t}}$

$$m_{\mathrm{I,bb}} = \frac{\sqrt{N_\nu}}{\mathrm{SNR}_{\mathrm{t}}}. \tag{3.13}$$

Throughout we will see that the number of frequency channels scales the modulation index but does not change the relative magnitude (i.e., signals with larger $f_\nu$ have lower $m_{\mathrm{I}}$). As the number of channels from a single instrument generally remains constant, it is unimportant to the interpretation of a single data set but is important when comparing data from different instruments and choosing an appropriate $m_{\mathrm{I,max}}$.

Equation 3.13 reveals a direct relationship between $\mathrm{SNR}_{\mathrm{t}}$ and $m_{\mathrm{I}}$. The modulation index of a broadband pulse is not arbitrary; rather, it must fall along a curve proportional to $1/\mathrm{SNR}_{\mathrm{t}}$. Furthermore for a signal with a given $\mathrm{SNR}_{\mathrm{t}}$, the broadband modulation index is the lowest $m_{\mathrm{I}}$ the signal may have on average, as any $f_\nu < 1$ increases the frequency modulation of a signal and correspondingly its modulation index. In Figure 3.1 the broadband pulses (closed circles) cluster in a stripe centered at $\mathrm{SNR}_{\mathrm{t}} = 10$ and $m_{\mathrm{I,bb}} = 1.6$. While each pulse has an inherent $\mathrm{SNR}_{\mathrm{t}} = 10$, the underlying noise in the data spreads the $\mathrm{SNR}_{\mathrm{t}}$ and $m_{\mathrm{I}}$ of individual realizations about the ensemble average value ($m_{\mathrm{I,bb}}$) and along the

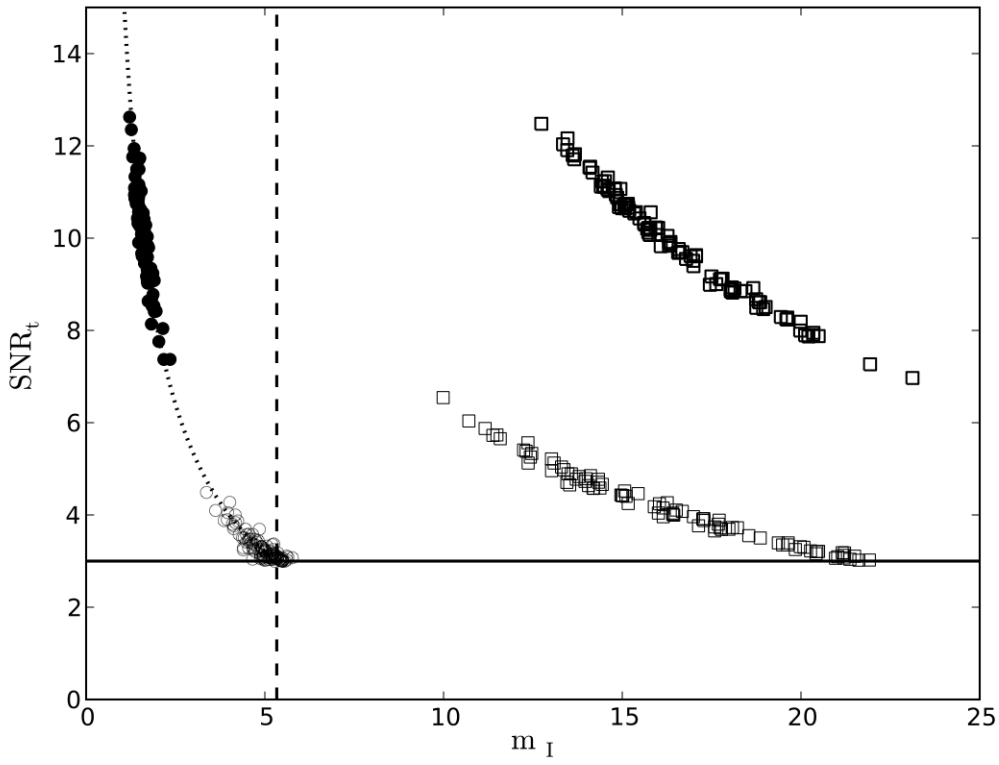


Figure 3.1: Time series SNR vs. modulation index for a simulated dataset with $N_\nu = 256$ containing noise (open circles), dispersed pulses (filled circles), and two sets of one-channel-wide RFI spikes (open squares). The solid horizontal line represents the applied intensity threshold of $\mathrm{SNR}_{\mathrm{min}} = 3$. The vertical dashed line represents $m_{\mathrm{I,T}} = 5.3$ as calculated for $\mathrm{SNR}_t = 3$ using Equation 3.14. The dotted curve shows $m_{\mathrm{I,bb}}$ as a function of $\mathrm{SNR}_t$ (Equation 3.13).

curve given by Equation 3.13 (dotted curve).

In practice there is a special $\mathrm{SNR_t}$: $\mathrm{SNR_{min}}$, the minimum SNR constraint applied to the time series in the "thresholding" operation. The corresponding modulation index is given by Equation 3.13

$$m_{\mathrm{I,T}} = \frac{\sqrt{N_\nu}}{\mathrm{SNR_{min}}}. \tag{3.14}$$

This modulation index represents the average maximum $m_{\mathrm{I,bb}}$ that may exist in a set of thresholded candidates. In real data there will be spectra with SNR above the $\mathrm{SNR_{min}}$ due to noise alone. The modulation index of thresholded noise follows the same relation as broadband pulses but with $\mathrm{SNR_t} \sim \mathrm{SNR_{min}}$. Most of the events due to noise will cluster near $m_{\mathrm{I,T}}$ and $\mathrm{SNR_{min}}$, but the rare stronger noise events will pepper the broadband curve toward larger $\mathrm{SNR_t}$ and lower $m_I$. In Figure 3.1 thresholded noise is shown as open circles, and only 100 points are plotted to reduce clutter. The weakest noise clusters near $\mathrm{SNR_{min}} = 3$ and $m_{\mathrm{I,T}} = 5.3$, and the stronger noise climbs the $m_{\mathrm{I}} \propto 1/\mathrm{SNR_t}$ curve (dotted curve).

Because $m_{\mathrm{I,T}}$ is on average the largest value of modulation index for a broadband signal in a candidate list, it is an upper limit to the choice of $m_{\mathrm{I,max}}$. Choosing a $m_{\mathrm{I,max}} > m_{\mathrm{I,T}}$ would only return events from thresholded noise or RFI. The dashed vertical line in Figure 3.1 shows $m_{\mathrm{I,T}}$ for $\mathrm{SNR_{min}} = 3$. If a modulation index upper limit is applied at $m_{\mathrm{I,max}} = m_{\mathrm{I,T}}$, events to the left of the line are kept (all of the broadband pulses and about 85% of the thresholded noise), while all events to the right are dropped (all the ultra-narrowband spikes and about 15% of noise events). A larger $\mathrm{SNR_{min}}$ would raise the solid line and move the dotted line to the left, which reduces the false alarm rate due to noise but limits one to detecting stronger pulses that are presumably rarer.

For the ultra-narrowband case where $f_\nu = 1/N_\nu$ and $\mathrm{SNR_t} \to \infty$, the depen-

dence on $SNR_t$ drops out, and the average modulation reduces to

$$m_{I,s} = \sqrt{N_\nu - 1}. \tag{3.15}$$

Note that the modulation index for spiky signals depends only on the number of channels. Equation 3.15 also gives the upper limit on $m_I$ as any $f_\nu > 1/N_\nu$ reduces the modulation and decreases the modulation index. Spiky RFI is illustrated in Figure 3.1 as two stripes of open squares centered at $SNR_t = 10$ and 5 respectively and $m_{I,s} = 16$. Like the broadband signals, the noise spreads the points about the ensemble average value, but clearly $m_{I,s}$ does not depend on $SNR_t$ for narrowband signals.

The modulation index for intermediate filling factors at constant $SNR_t$ must transition smoothly from $m_{I,bb}$ to $m_{I,s}$ as $f_\nu$ goes from 1 to $1/N_\nu$. The exact manner of the transition is given by Equation 3.12. Figure 3.2 illustrates the analytical variation of modulation index with filling factor for four values of the time series SNR ranging from $SNR_t = 3$ (top) to $SNR_t = 100$ (bottom) and $N_\nu = 256$. At low filling factors all curves tend toward $\sqrt{N_\nu}$, and at high filling factors the stronger the signal, the lower the modulation index. Most of the drop in modulation index happens at $f_\nu < 0.1$, suggesting that this technique can easily classify signals with filling factors less than about 10% but is less sensitive for signals with moderate to high filling factors.

In the above analysis we have assumed that a signal is either narrowband or broadband. Reality is messier, and we might have overlapping signals that are both narrowband and broadband. For example a pulse might occur at the same time as persistent narrowband RFI. To explore this case we adopt a simple model looking at the modulation index of a broadband pulse with no intrinsic frequency structure contaminated by an RFI spike that is one channel wide

(i.e. the "ultra-narroband" case described above). Equation 3.16 estimates the modulation index for this two-component case:

$$m_{\mathrm{I,bb+s}} = \frac{\sqrt{N_\nu(1 + \mathrm{SNR}_{\mathrm{t,s}}^2)}}{\mathrm{SNR}_{\mathrm{t,bb}} + \mathrm{SNR}_{\mathrm{t,s}}}, \tag{3.16}$$

where $\mathrm{SNR}_{\mathrm{t,bb}}$ and $\mathrm{SNR}_{\mathrm{t,s}}$ are the time series SNR of the broadband pulse and RFI spike respectively. Note when $\mathrm{SNR}_{\mathrm{t,s}} \to 0$, the above equation reduces to the expression for broadband pulses (Equation 3.13), and when $\mathrm{SNR}_{\mathrm{t,bb}} \to 0$, the above equation reduces to the expression for ultra-narrowband spikes (Equation 3.15, up to the "$-1$", which is negligible for large $N_\nu$). When $\mathrm{SNR}_{\mathrm{t,bb}} \sim \mathrm{SNR}_{\mathrm{t,s}}$, we see that $m_{\mathrm{I,bb+s}} \approx 0.5 m_{\mathrm{I,s}}$. If we choose the modulation index cutoff to be $m_{\mathrm{I,T}}$, this implies a $\mathrm{SNR}_{\mathrm{min}} \leq 2$ to allow $m_{\mathrm{I,bb+s}}$ to be above threshold, which is an unreasonably low threshold for most surveys. Furthermore using Equation 3.16, we can estimate that in order for $m_{\mathrm{I,bb+s}} \geq m_{\mathrm{I,T}}$, $\mathrm{SNR}_{\mathrm{t,bb}}/\mathrm{SNR}_{\mathrm{t,s}} > \mathrm{SNR}_{\mathrm{min}} - 1$. For example if $\mathrm{SNR}_{\mathrm{min}} = 5$, the $\mathrm{SNR}_{\mathrm{t}}$ of the pulse must be four times larger than the $\mathrm{SNR}_{\mathrm{t}}$ of the narrowband RFI. Using only the $\mathrm{SNR}_{\mathrm{t}}$ and modulation index, we could easily miss a pulse if it occurs concurrent with strong, narrowband RFI. This finding that the modulation index is more sensitive to narrowband signals than broadband signals is consistent with Figure 3.2 and suggests our method is not a substitute for RFI excision techniques that identify persistent, strong, narrowband RFI from raw data.

### 3.2.4 Interstellar scintillations

Small scale density irregularities in the ionized interstellar medium (ISM) scatter and refract radio waves. Diffractive interstellar scintillations (DISS) and refractive interstellar scintillations (RISS) are observational phenomena seen in

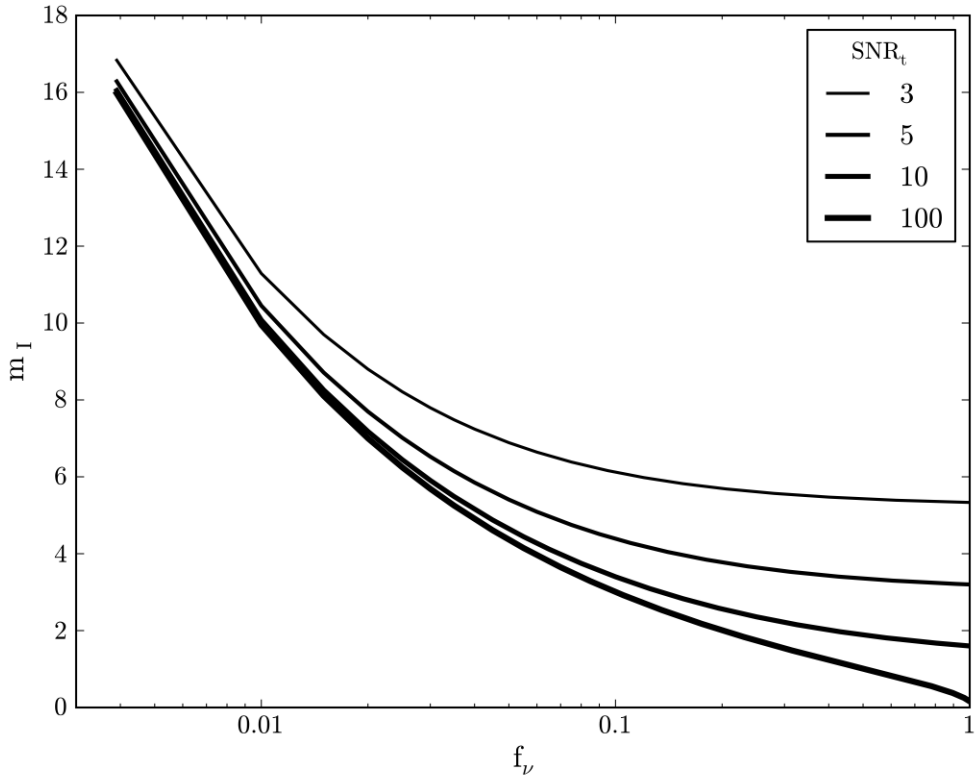


Figure 3.2: Modulation index vs. filling factor calculated using Equation 3.12 with $N_\nu = 256$. From top to bottom (thin to thick), curves are shown for four values of $SNR_t$: 3, 5, 10, 100.

compact radio sources due to these irregularities. Scintillations are characterized by intensity variations with typical time ($\Delta t_{\rm DISS}$) and frequency scales ($\Delta \nu_{\rm DISS}$), and for DISS in the strong scattering regime, a time series is modulated as a random variable $g_{\rm DISS}(t)$ with an exponential amplitude distribution and fractional intensity variations on the order of unity. More generally, DISS varies with both time and frequency ($g_{\rm DISS}(t, \nu)$) with the diffraction timescale scaling as $\Delta t_{\rm DISS} \sim \nu^{1.2}$ and bandwidth scaling as $\Delta \nu_{\rm DISS} \sim \nu^{4.4}$. Also note that $\Delta \nu_{\rm DISS}$ and the pulse broadening time ($\tau_{\rm d}$) are Fourier transform pairs and given by

$2\pi\Delta\nu_{\rm DISS}\tau_{\rm d} = C_1$ (Cordes & Rickett, 1998) where $C_1 \approx 1$. For fast transients we can safely assume that the duration of the pulse is much less than the diffraction time scale (at least for DM $\lesssim$ a few hundreds at $\nu = 1$ GHz), whereas the scintillation bandwidth may be of the same order as the channel resolution or bandwidth of a spectrometer. Frequency structure caused by DISS will increase the variance and thereby increase the modulation index of a signal.

Understanding the frequency structure of a spectrum due to strong scattering is clarified by defining three limits based on the relative sizes of $\Delta\nu$, $B$, and $\Delta\nu_{\rm DISS}$ where $\Delta\nu$ is the width of a single frequency channel in a spectrum (Cordes et al., 2004). When the scintillation bandwidth is smaller than the channel bandwidth ($\Delta\nu_{\rm DISS} \ll \Delta\nu$), the instrument effectively averages over several "scintles" and the intensity modulations are quenched. In this case $m_{\rm A} \ll 1$, and the modulation index follows the expressions in Section 3.2.3. Similarly in the other extreme where there is only one scintle across the band ($0.2B < \Delta\nu_{\rm DISS}$), the intensity variation is nearly flat over the entire bandwidth and $m_{\rm A} \ll 1$. For the intermediate case where there are several distinct scintles across the band ($\Delta\nu \lesssim \Delta\nu_{\rm DISS} \lesssim 0.2B$), the amplitude of each scintle is exponentially distributed and $m_{\rm A} \approx 1$. As instruments become increasingly wide band and $B/\nu \sim 1$, this situation will become ever more relevant. Examples of these three cases can be seen in Cordes et al. (2004) for giant pulses from the Crab pulsar.

Connecting this to the discussion in Section 3.2.3, the resulting average modulation index of a spectrum containing Gaussian noise and a broadband signal with exponentially distributed amplitudes is given by Equation 3.12 with $f_\nu = 1$ and $m_{\rm A} = 1$. For weak signals the leading term of Equation 3.12 dominates and the spectral modulation index increases only slightly over that for a perfectly

flat signal (i.e. Equation 3.13). For example, the modulation index of a spectrum with $N_\nu = 256$ and $\mathrm{SNR}_{\mathrm{min}} = 5$ increases from $m_\mathrm{I} = 3.2$ to $m_\mathrm{I} = 3.35$ . For strong signals (i.e. $\mathrm{SNR}_\mathrm{t} \gg \sqrt{N_\nu}$) the first term in Equation 3.12 becomes negligible and $m_\mathrm{I} \approx m_\mathrm{A} \approx 1$ for all $\mathrm{SNR}_\mathrm{t} \gg \sqrt{N_\nu}$.

### 3.2.5 Self Noise in the Pulsar Signal

Broadband pulsar signals have been modeled as amplitude modulated noise (Rickett et al., 1975) and as modulated, polarized shot noise (Cordes, 1976; Cordes et al., 2004). The noise in these models corresponds to the emission over a broad range of radio frequencies while the modulation accounts for pulse structure. In this context, the modulation index of the signal is nonzero but is still smaller than the modulation indices expected from RFI. In the limit where the noise has Gaussian statistics, the intensity modulation index of polarized noise is

$$m_I^2 = m_\mathrm{ISS}^2 + (1 + m_\mathrm{ISS}^2)(1 + d_p^2)/2, \tag{3.17}$$

where $m_\mathrm{ISS}$ is the modulation index of frequency structure from DISS and $d_p$ is the degree of polarization. The largest modulation is for $m_\mathrm{ISS} = d_p = 1$ when $m_I = \sqrt{3}$. The observed modulation will be reduced if frequency structure from DISS is much broader than the total bandwidth $B$ or if multiple pulse structures are averaged over in a single frequency channel of the spectrometer.

### 3.2.6 Signature of Incorrect Dedispersion

A bright pulse in a survey dedispersed with a large number of trial dispersion measures will yield events at many incorrect DMs in addition to the correct one. The true DM will return the largest $\mathrm{SNR_t}$ and narrowest pulse width on average. The residual pulse smearing from neighboring, incorrect DMs yields a lower $\mathrm{SNR_t}$ and wider pulse. The larger the DM error, the smaller the SNR is on average until the SNR drops below the threshold. This SNR–DM signature is one of the tests that a signal is a true astrophysical pulse and not RFI. The modulation index of a pulse also increases as the DM error increases.

The two-dimensional filling factor for a dispersed pulse dedispersed with a DM error of $\delta$DM is

$$f_{\nu,t} = \left[1 + \frac{|\Delta t_{\delta\mathrm{DM}}|}{\Delta t_\mathrm{p}}\right]^{-1}, \tag{3.18}$$

where $|\Delta t_{\delta\mathrm{DM}}|$ is the absolute value of the residual dispersion smearing and $\Delta t_\mathrm{p}$ is the intrinsic width of the pulse in seconds. For a pulse that is perfectly dedispersed, $\Delta t_{\delta\mathrm{DM}} = 0$ and $f_{\nu,t} = 1$, while for a large DM error, $\Delta t_{\delta\mathrm{DM}} \to \infty$ and $f_{\nu,t} \to 0$.

Ideal matched filtering of a dispersed pulse, either in the absence of dedispersion or from residual smearing from incorrect dedispersion, smooths over the duration of the pulse's dispersion sweep, consolidating the signal into a single time bin. The signal has a lower $\mathrm{SNR_t}$ than the pulse's intrinsic $\mathrm{SNR_t}$ because $f_{\nu,t} < 1$, but the one-dimensional filling factor is still $f_\nu = 1$. The spectral modulation index increases slightly due to the drop in $\mathrm{SNR_t}$ (Equation 3.13). For example, a hypothetical, unresolved Crab giant pulse detected in the data set described in Section 3.3.4 with an intrinsic $\mathrm{SNR_t} = 30$ has $m_\mathrm{I} = 0.75$ when dedispersed using the true DM of the Crab pulsar (DM = 56.71 pc cm$^{-3}$) and $m_\mathrm{I} \approx 4.3$

when dedispersed with a DM error of $\delta$DM = 2 pc cm$^{-3}$.

The spectral modulation index does reflect the signature of incorrect dedispersion but only indirectly. A more powerful statistic would depend on $f_{\nu,t}$ rather than $f_\nu$. This is accomplished by calculating an alternative modulation index ($m_{I,r}$) using the time resolved intensity moments given by Equation 3.9. For the same hypothetical Crab giant pulse described above, the resolved modulation index is the same as the spectral modulation index ($m_\mathrm{I}$ = 0.75) when the pulse is correctly dedispersed, but it increases to $m_{I,r}$ = 24.7 for a DM error of $\delta$DM = 2 pc cm$^{-3}$.

Combining the information provided by the spectral modulation index and resolved modulation index helps to identify spurious events from incorrect dedispersion. By first applying a cutoff in spectral modulation index, events are classified as either broadband or narrowband in frequency. The resolved modulation index of the broadband signals distinguish those that are also broad in time from those that are narrow in time.

### 3.2.7 Modulation Index Cutoff

The role of the spectral modulation index in a source detection pipeline is analogous to the SNR threshold (SNR$_{\mathrm{min}}$). A signal is first classified as interesting or not based on its SNR. For candidate signals that are strong enough, the application of a spectral modulation index cutoff ($m_{\mathrm{I,max}}$) classifies the candidate as interesting or not based on how broadband it is. Just as SNR$_{\mathrm{min}}$ is applied to data automatically, the modulation index cutoff can be applied without human supervision.

The choice of $m_{\mathrm{I,max}}$ is influenced by the characteristics of the instrumentation and pipeline, as well as the importance of astrophysical effects such as diffractive interstellar scintillations. The upper limit to $m_{\mathrm{I,max}}$ is given by modulation index at the SNR threshold ($m_{\mathrm{I,T}}$). As explained in Section 3.2.3, $m_{\mathrm{I,T}}$ is the largest modulation index on average for a broadband pulse with a perfectly flat spectrum in a survey with an applied SNR minimum $\mathrm{SNR_{min}}$ (Equation 3.14). Choosing $m_{\mathrm{I,max}} > m_{\mathrm{I,T}}$ only yields candidates from thresholded noise or narrowband RFI. One exception is astrophysical signals with frequency structure, such as that caused by DISS. But as we showed in Section 3.2.4, the increase in the modulation index for weak signals is small (~ a few percent), and one could increase $m_{\mathrm{I,max}}$ to allow for weak, scintillating signals with minimal increase in false positives.

Choosing $m_{\mathrm{I,max}} < m_{\mathrm{I,T}}$ will reduce the number of false positives caused by thresholded noise and incorrectly dedispersed pulses, because it effectively applies a larger $\mathrm{SNR_{min}}$ (Equation 3.13). The cost is reduced sensitivity, as on average only pulses with $\mathrm{SNR_t}$ exceeding than the larger effective $\mathrm{SNR_{min}}$ will be below $m_{\mathrm{I,max}}$. But for surveys where interstellar scintillation may be important, a hard lower limit to the modulation index cutoff is set by the nature of exponential statistics; namely, $m_{\mathrm{I,max}} = 1$, as described in Section 3.2.4.

Real signals will not always cleanly divide into broadband or narrowband. For example RFI could be marginally broadband, a real pulse could have strong, narrowband structure, or both narrowband and broadband signals could occur simultaneously. But for fast transients we also have an additional parameter: dispersion. Looking at the peak dispersion measure of a candidate signal can break the degeneracy between RFI and an astrophysical signal with the same

modulation index.

### 3.2.8 Correlation Bandwidth

While the modulation index provides information about the degree of modulation in a spectrum, it does not provide any information about the shape of the modulation. A signal with amplitude $\overline{A}$ localized in $N$ adjacent bins has the same variance as a signal with the same amplitude whose power is distributed in $N$ isolated bins across the band. We therefore need a new measure that reflects the distribution of a signal in frequency.

We define the characteristic bandwidth ($B_c$) to be the typical width of the frequency structure of signal. A broadband signal has $B_c = B$, and a narrowband signal has $B_c \sim \Delta\nu$. We define the fractional correlation bandwidth as a measure of the typical correlation scale compared to the total bandwidth

$$\mathrm{FCB} = \frac{\mathrm{B_c}}{\mathrm{B}}. \tag{3.19}$$

The $B_c$ can be estimated by calculating the correlation length from the autocorrelation (ACF) of a spectrum, which we define to be the half width half maximum (HWHM) of the first lobe of the ACF.

Figure 3.3 illustrates the technique for two simulated spectra with the same total intensity and variance. The left spectrum contains a Gaussian-shaped signal with a FWHM of 32 frequency channels. The right spectrum was generated by randomly swapping the channels in the left spectrum in groups of four, so that both spectra have the same total intensity, variance, and modulation index ($m_I \approx 1.5$). The bottom panels show the ACFs for the two spectra. To have the

zero lag of the ACF equal to $m_I^2$, the ACF was scaled and offset by ACF/$(N_\nu \bar{I}^2) - 1$, where $\bar{I}$ is the mean of the spectrum. For the Gaussian spectrum, FCB = 0.18 and is consistent with the autocorrelation of a Gaussian with a FWHM=32 channels. For the spiky spectrum, FCB = 0.03 is consistent with a signal with spikes approximately four channels wide. Both spectra could be interesting astrophysical signals. The left signal could clearly be caused by an astrophysical process. The right signal could be indicative of strong interstellar scintillations. In any case the fractional correlation bandwidth provides another parameter that can be used to automatically classify a signal. Note the values for the fractional correlation bandwidth were calculated automatically along with the ACF and other statistics, suggesting the FCB could be implemented in an unsupervised pipeline.

## 3.3 Application

This section applies the techniques described above to the detection of fast radio transients. First we discuss one approach of incorporating the calculation of the spectral modulation index into a transient detection pipeline. To illustrate the technique we simulate a transient detection pipeline and apply the detection scheme to real data containing known transients. Our simulations and applications to real data show that a modulation index cutoff efficiently flags RFI and significantly reduces the number of candidates. Note our implementation starts with a list of pulse candidates generated in the standard manner and not on the raw data directly. A discussion of how the modulation index could be used to flag raw data in real time is discussed in Section 3.4.4. Also note our technique

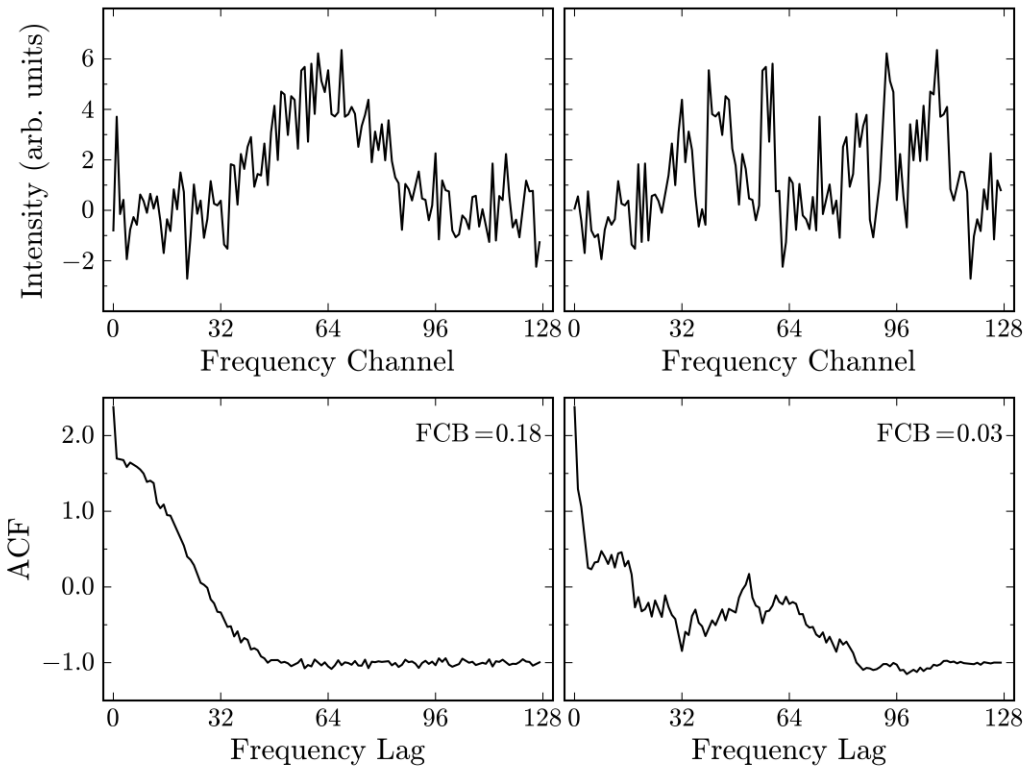


Figure 3.3: Fractional correlation bandwidth (FCB) for two example spectra. The top panels show two simulated spectra with the same total intensity and variance and $N_\nu = 128$. The left spectrum contains a Gaussian with a FWHM frequency width of 32 channels, and the right spectrum is a spiky spectrum generated by randomly swapping groups of four channels in the left spectrum. The autocorrelation functions (ACF) of the sample spectra are shown in the bottom panels. The ACF has been normalized by the square of the mean of the spectrum and the number of channels all minus 1. The characteristic bandwidth is calculated as the half width at half maximum of the first lobe of the ACF, and the corresponding FCB is given in the upper right corner of the lower panels.

is independent of the number of beams (e.g. Arecibo L-band Feed Array[2]) or stations (e.g. Very Long Baseline Array, Thompson et al., 2011) used to collect the data.

### 3.3.1 Implementation in a Detection Pipeline

The pipeline makes two passes over the data. The first pass defines a list of candidate events, and the second pass calculates the second moment and modulation index of these events. This two-pass approach is adopted for practicality and flexibility and will be explained in detail below. Also recall that the modulation index requires the first and second *central* moments, so the data must be bandpass subtracted before $\bar{I}$ and $\overline{I^2}$ are calculated.

The data are first dedispersed as described by Equation 3.3 to generate a first moment time series. Surveys use a range of trial dispersion measures, each generating its own time series. These dedispersed time series are smoothed in time by applying a template bank of matched filters with different properties to account for pulses of different widths and profiles. A list of candidates is defined by applying a minimum SNR threshold ($\mathrm{SNR}_{\mathrm{min}}$) to these smoothed, dedispersed time series. In this paper we focus on two specific implementations of matched filtering: boxcar smoothing and clustering (friends-of-friends).

Boxcar smoothing convolves the data with a boxcar function of length $W_t$ samples. The corresponding smoothing weights for Equation 3.5 are $w_{tt'} = 1$ for $t' = 0, ..., W_t$. In practice boxcar matched filtering is implemented by iteratively summing adjacent time samples so $W_t = 2^{n_{sm}}$ where $n_{sm}$ is the number of

[2] http://www.naic.edu/alfa

smoothing iterations. The details of this technique are described in Cordes & McLaughlin (2003). A single signal will likely be detected at several values of $n_{sm}$, so the event list should be sifted for the boxcar width that yields the maximum SNR.

The cluster, or friends-of-friends, algorithm looks for groupings of events in a time series. A cluster is defined as a set of events for which there is no gap in samples larger than $N_{gap}$. For a cluster with $N_{cluster}$ samples, the smoothing weights are $w_{tt'} = 1$ for $t' = t_1, ..., t_N$ where $t_1$ is the first sample in a cluster, $t_N$ is the last sample, and $t'_{i+1} - t'_i < N_{gap} + 1$. Importantly, the cluster algorithm is agnostic to the symmetry of the pulse, unlike the boxcar smoothing filter which is symmetric. As pulses from highly scattered sources have exponential tails, this algorithm may be better suited to detecting such astrophysical objects.

After a list of possible candidates is defined, one goes back to the raw time-frequency data to calculate the second moments and modulation indices. We grab a narrow range of raw data centered at the location of the candidate and reprocess it. The time-frequency snapshot is dedispersed and smoothed at the DM and smoothing parameters determined from the first pass. A time series is calculated for both the first and second moments. The first moment time series is thresholded in intensity, and the modulation index is calculated for the samples above threshold. The modulation indices of the events are compared to the modulation index cutoff, $m_{\mathrm{I,max}}$, and flagged as either a signal of interest if $m_I \leq m_{\mathrm{I,max}}$ or RFI if $m_I > m_{\mathrm{I,max}}$.

The two-pass approach is adopted out of practical considerations. While the first moment of the dedispersed time series can be smoothed directly, the time-frequency data must be smoothed before being squared and averaged in fre-

quency (Equation 3.7). This would require a different dedispersed, smoothed, second moment time series for each matched filter type and parameter. Furthermore some smoothing approaches, such as the cluster algorithm, determine the smoothing weights $w_{tt'}$ from the thresholded first moment time series, making a parallel calculation of the second moment time series impractical.

### 3.3.2 Processing Requirements

For most cases the two pass approach is more computationally efficient than the obvious alternative of calculating the second moment in parallel with the first. We parameterize the processing required by the second pass in terms of the processing required to do the dedispersion in the first pass. Generally dedispersion dominates the processing time in a transient survey, so this is a useful metric. The number of operations required to dedisperse a block of time-frequency data with $N_t$ time samples and $N_\nu$ frequency channels with $N_{DM}$ trial dispersion measures is $N_{ops,1} = N_t \times N_\nu \times N_{DM}$. If our first pass generates $N_{events}$ candidate events, the number of operations required to dedisperse a narrow time range with $N_s$ samples around each event at a single DM is $N_{ops,2} = 4N_{events} \times N_s \times N_\nu$. The factor of 4 was included to consider the squaring of the data, summing of both the original and squared data and bandpass subtraction. The processing required by the second pass normalized by the dedispersion processing of the first pass is

$$P_2 = \frac{N_{ops,2}}{N_{ops,1}} = \frac{4N_{events}N_s}{N_{DM}N_t}. \tag{3.20}$$

If $N_t = 10^6$, $N_{DM} = 10^3$, $N_s = 10^3$, $N_{events} = 10^4$, $P_2$ is a few percent of the original dedispersion processing. This analysis ignores the additional overhead incurred

by returning to the raw data, and in particular one must be wary of excess file I/O.

Most RFI excision techniques operate on raw data, not on the list of candidate pulses. Our technique is more general, because it can be used at both the beginning (see Section 3.4.4) and end of a source detection pipeline. The most common RFI excision approach applied to a list of candidates is to remove all events at low dispersion measure under the assumption that terrestrial signals are not dispersed. This is a blunt instrument and does not remove events from RFI at higher DMs. Calculating the modulation index of the candidate events allows for more sophisticated RFI excision in a list of candidates.

Classifying signals with the modulation index should be used in conjunction with other RFI excision algorithms. As described in Section 3.2.3, the modulation index is more sensitive to narrowband signals, and weak pulses may be missed if they occur simultaneously with strong, narrowband RFI. This suggests the modulation index algorithm works best together with algorithms that remove channels that contain persistent RFI. Similarly broadband RFI has a low modulation index, so removing impulsive RFI through other means will reduce the number of events from RFI that fall below the modulation index cutoff.

### 3.3.3 Simulations

To assess the usefulness of the modulation index as a signal diagnostic, we simulated a single-pulse event detection pipeline using Python-based software. In our simulations fake time-frequency data can be generated with Gaussian-distributed noise, dispersed pulses, and a variety of RFI. A dispersed pulse is

added with a specified dispersion measure and Gaussian pulse profile with a FWHM of $W_t$. Spiky RFI is modeled as a two-dimensional Gaussian with a FWHM width in time ($W_\mathrm{t}$) and frequency ($W_\nu$). Broadband RFI is modeled as an undispersed pulse (i.e., DM=0 pc cm$^{-3}$). The simulated data then undergo single pulse search processing. First the time-frequency data are dedispersed over a range of trial dispersion measures and both $\bar{I}$ and $\overline{I^2}$ is calculated from the dedispersed time series. $\bar{I}$ is thresholded, and for samples that are above threshold, the modulation index is calculated. Note that because these simulations involve a small amount of data, the two-pass analysis as described in Section 3.3.1 is not necessary, because all intermediate data products (i.e., dedispersed time series) can be kept in memory.

The fake data shown in panel (a) of Figure 3.4 set contains a single dispersed pulse, a Gaussian RFI spike, and a broadband RFI spike. The data properties are $N_\nu = 256$, $\Delta t_\mathrm{s} = 1$ ms, $N_\mathrm{t} = 2000$ time samples, $\nu_o = 1400$ MHz, and $B = 100$ MHz. A single dispersed pulse was added at $t = 0.25$ s with DM $= 500$ pc cm$^{-3}$, SNR$_{\nu\mathrm{t}} = 1$, and $W_\mathrm{t} = 1$. A narrowband Gaussian spike was added at $t = 1.0$ s and $\nu_o \approx 1428$ MHz with SNR$_{\nu\mathrm{t}} = 40$, $W_\mathrm{t} = 2$, and $W_\nu = 2$. Finally a broadband RFI spike is represented by an undispersed pulse (i.e., DM = 0 pc cm$^{-3}$) at $t = 1.5$ s with SNR$_{\nu\mathrm{t}} = 5$ and $W_\mathrm{t} = 1$. The data were dedispersed over a range of trial dispersion measures DM = 0 − 1000 pc cm$^{-3}$ and DM interval $\Delta$DM = 6 pc cm$^{-3}$. A list of candidate event was defined by applying an SNR threshold of SNR$_{\mathrm{min}} = 3$ to the resulting time series.

Figure 3.4 shows the results of this simulation. The top frame shows (a) the time-frequency data, (b) candidate events vs. DM and time, (c) $50/m_\mathrm{I}$ for the candidate events, and (d) intensity vs. DM and time for candidates below

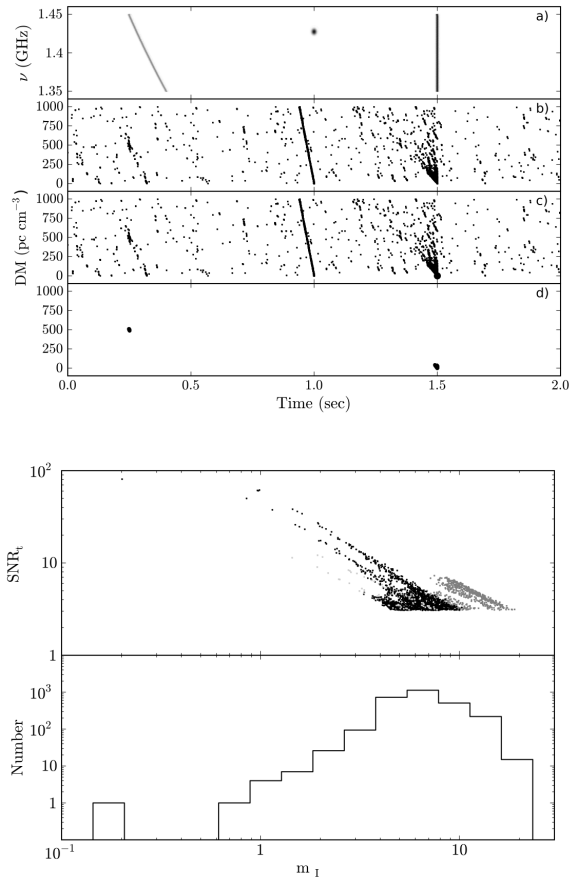


Figure 3.4: Diagnostic plots for the simulated data described in Section 3.3.3. Top Frame: Panel (a) shows a grey scale of the simulated time-frequency data where the SNRs and widths of the signals have been exaggerated for clarity. Panel (b) plots a point in the DM-time plane for each event above the intensity threshold ($\mathrm{SNR_{min}} = 3$). Panel (c) plots $50/m_I$ for the candidate events in panel (b). Panel (d) plots the SNR of candidates over the SNR threshold and below the modulation index threshold ($m_{\mathrm{I,max}} = 3.2$). The data were dedispersed over a DM range of 0 to 1000 pc cm$^{-3}$ with $\Delta$DM = 6 pc cm$^{-3}$. The dispersed pulse is located at DM = 500 pc cm$^{-3}$ and at $t = 0.5$ s. A Gaussian RFI spike is located at $t = 1.0$ s and a frequency of $\sim 1428$ MHz and causes the stripe of events between $t \approx 0.9 - 1.0$ s. Broadband RFI is located at $t = 1.5$ s and causes the triangle-shaped patch of candidate events. Bottom Frame: The top panel plots $\mathrm{SNR_t}$ vs. $m_{\mathrm{I}}$, and the bottom panel is a histogram of modulation index. Events associated with the broadband RFI (and thresholded noise) are plotted in black, events associated with the Gaussian RFI spike are shown in medium gray, and the events associated with the dispersed pulse are plotted in light gray.

$m_{\mathrm{I,max}} = 3.2$. For clarity the SNR and pulse widths of the three signals in panel (a) have been exaggerated, and the noise is not shown. Instead of plotting the intensity, panel (b) shows a dot for each sample that was above the intensity threshold to reduce clutter. Panel (c) shows that most of the events have similar modulation indices; the one exception is for the broadband pulse, which has a modulation index almost an order of magnitude lower than all other points due to its high SNR. Panel (d) plots the SNR of the events with modulation indices below $m_{\mathrm{I,max}}$ with the area of the circle proportional to the SNR of the event.

The triangle-shaped group of events near $t \sim 1.5$ s in panels (b) and (c) are spurious hits caused by the dedispersion path crossing the bright, broadband RFI samples. Most of these events have low filling factors because the RFI samples contribute only a few samples to the dedispersed spectrum and are not present in panel (d). The exception is low DM and $t = 1.5$ s where signal from the RFI contributes to many frequency channels resulting in a low modulation index. This is the incorrect dedispersion effect described in Section 3.2.6. While this example of RFI does pass our modulation index filter, the low DM of the event exposes it as RFI. Applying yet another filter that removes events at low DM will remove such events. The narrowband Gaussian spike causes a stripe of spurious events between $t \approx 0.9 - 1.0$ s as the path of the each trial dispersion measure crosses the spike. But for reasons just described above, these spurious candidates have modulation indices above our threshold. Finally the true astrophysical dispersed pulse, which is buried in the middle panels of Figure 3.4, stands out in panel (d) at $t = 0.25$ s after applying the modulation index cutoff.

The bottom frame in Figure 3.4 plots $\mathrm{SNR_t}$ vs. $m_{\mathrm{I}}$ and a histogram of $m_{\mathrm{I}}$ to illustrate how the modulation index groups signal types. The spurious events

caused by the Gaussian spike (medium gray) are clumped together on the far right of the plot between $8 < m_I < 20$, consistent with the extreme narrowband case given by Equation 3.15 ($m_{I,s} = 16$). The events due to the dispersed pulse and broadband RFI follow the light gray and black tracks respectively. The tracks are caused by incorrectly dedispersing the signal, and in both cases the lowest modulation index corresponds to the correct dispersion measure. The events from thresholded noise (black) are clumped near $m_{I,T} \sim 5$ and $SNR_t \sim SNR_{min}$ as predicted by Equation 3.14. Because the weaker events associated with the broadband RFI overlap with the area containing thresholded noise, we've plotted them with the same color. The histogram of $m_I$ in the lower panel shows that by applying a $m_{I,max} = 3.2$, we have eliminated most of the events.

### 3.3.4 Application to Data

In the next two subsections we apply our modulation index classification technique to two known sources of single pulses detectable by single-pulse search pipelines. In Section 3.3.4 we apply the method to RRAT J1928+15 and in Section 3.3.4 to giant pulses from the Crab pulsar. Although in both cases we know the correct dispersion measure of the source, we re-analyze the data at a range of trial dispersion measures to recreate typical survey results.

The data were bandpass corrected by dividing by the median bandpass and subtracting off the mean. Dividing by the median spectrum flattens the spectra, assuring the bandpass shape is not contributing to the variance calculation. We choose the median because it is less sensitive to extreme values caused by pulses or RFI. Subtracting the mean assures us that our spectra have zero mean.

**RRATs**

RRAT J1928+15 was discovered by Deneva et al. (2009) in the PALFA survey running at the Arecibo Observatory. (See Deneva et al. 2009 for the details of the observation and PALFA parameters.) In brief, J1928+15 was discovered using the single-pulse search algorithms implemented by the Cornell pulsar search pipeline. Three pulses were detected with an interval of 0.403 s at DM = 242 pc cm$^{-3}$. We reprocessed the bandpass corrected data with 642 trial dispersion measures roughly equally spaced over a DM range DM = 0 – 500 pc cm$^{-3}$. The dedispersed time series had an intensity threshold applied at $SNR_{min} = 4$. The samples above the SNR threshold were run through the cluster algorithm, and the sample in each cluster with the maximum intensity was used to calculate the modulation index.

The results are shown in Figure 3.5. In all panels only events with SNR > 5 are shown. The top frame plots intensity versus DM and time for all events above threshold (top) and the events that also satisfy $m_I < 1$. (Compare to Figure 3 in Deneva et al. 2009). The strong RRAT pulse is clearly visible around t = 100 s, and a weak broadband RFI spike occurs near $t = 89$ s. For the strong RRAT pulse we measure $SNR_t \approx 20$, and for the preceding weaker pulse, $SNR_t \approx 6$. The difference between our values and those in Deneva et al. (2009) is likely due to the fact that we used the cluster algorithm and they used the boxcar smoothing matched filter algorithm. We also do not detect the weaker, tailing pulse, likely for the same reason. Applying a $m_{I,max} = 1$ eliminates $\approx$ 99% of the events, leaving only the strongest events associated with the bright RRAT pulse.

The bottom frame shows $SNR_t$ vs. $m_I$ (top) and a histogram of $m_I$ (bottom). In the top panel the events associated with the strong RRAT pulse are localized

along the upper curve extending from $\mathrm{SNR_t} \approx 20$ and $m_\mathrm{I} \approx 0.8$ to $\mathrm{SNR_t} \approx 5$ and $m_\mathrm{I} \approx 3$. The lower limit of $m_\mathrm{I} \approx 0.8$ is consistent with the value predicted by Equation 3.13 for $\mathrm{SNR_t} = 20$. The curve itself comprises spurious events caused by the strong pulse being dedispersed at incorrect dispersion measures and account for about 90% of all events in this figure. As described in Section 3.2.6, applying a modulation index cutoff can reduce the number of spurious events from real signals at incorrect DMs. The broadband RFI spike produced the events along the vertical line along $\mathrm{SNR_t} = 5$ and $1.25 < m_\mathrm{I} < 2.5$. The events associated with thresholded noise are grouped between $2.5 < m_I < 3.5$ and are consistent with $m_\mathrm{I,T} = 3.2$ for $N_\nu = 256$ and $\mathrm{SNR_{min}} = 5$. While applying $m_\mathrm{I,max} = 1$ cleanly isolated the bright pulse from the RRAT, we did so at the cost of applying an effective $\mathrm{SNR_{min}} = 10$ and thereby removed the weaker RRAT pulse.

**Crab giant pulses**

The most well-known source of single, dispersed pulses is the Crab pulsar. We reprocessed a 430 MHz data set containing the "supergiant" pulse with a $\mathrm{SNR_t} \sim 1500$ reported in Cordes et al. (2004). The 140-second long dataset was dedispersed using 616 trial dispersion measures ranging from $\mathrm{DM} = 0 - 250\,\mathrm{pc\ cm^{-3}}$. An intensity threshold of $\mathrm{SNR_{min}} = 4$ was applied to the dedispersed time series, and the samples above threshold were run through the cluster algorithm. The modulation index was calculated for the sample from each cluster that had the maximum intensity. This data set proved to be a particularly interesting case study due to the large number of pulses with a range of SNRs.

The results of the reprocessing are shown in Figure 3.6, and in all panels

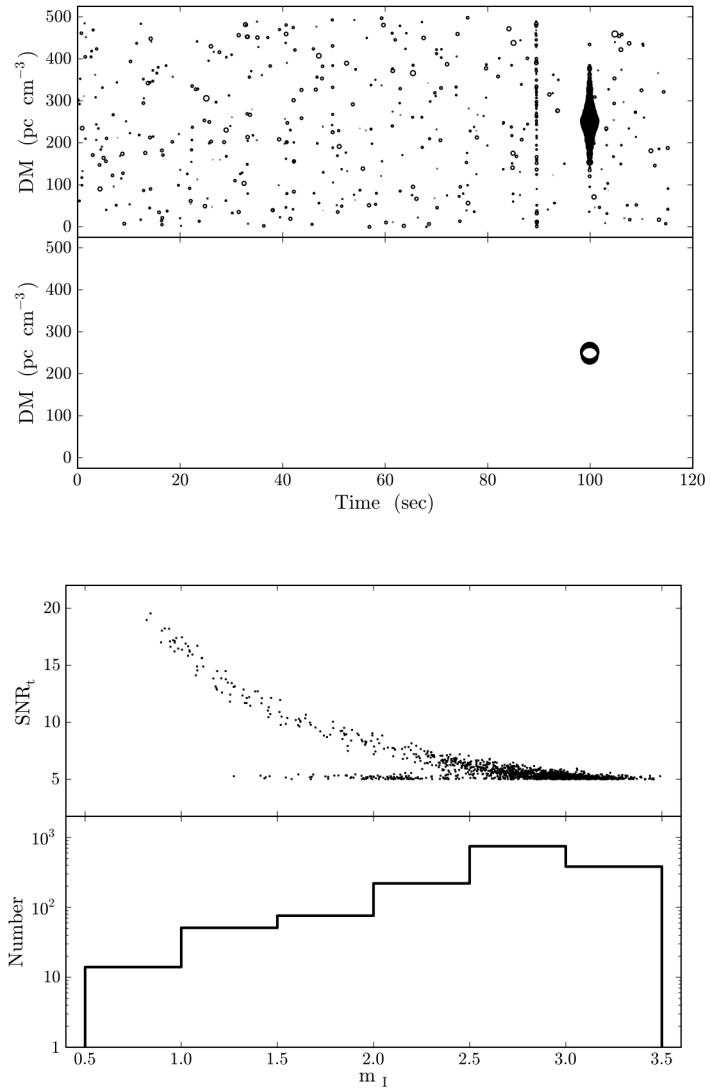


Figure 3.5: Modulation index calculation for the RRAT J1928+15. In all frames events with SNR > 5 are shown. Top Frame: Intensity vs. DM and time (top) for the events above the SNR threshold and the same events (bottom) with an additional modulation index threshold ($m_{\mathrm{I,max}} < 1$) applied (bottom). The area of each circle is proportional to the SNR of the event. (See Deneva et al. 2009, Figure 3.) The RRAT pulse is clearly visible around $t$ = 100 s and DM = 242 pc cm$^{-3}$ and a weak, broadband RFI spike around $t$ = 89 s. Bottom Frame: The SNR$_t$ vs. $m_{\mathrm{I}}$ (top) and a histogram of $m_{\mathrm{I}}$ (bottom) . The data were dedispersed using $\sim$ 640 trial dispersion measures ranging from 0 to 500 pc cm$^{-3}$.

only events with SNR > 5 are shown. The top frame shows the intensity for the events versus DM and time with a circle whose area is proportional to the SNR of each event. The top panel shows all events, and the bottom panel shows those events that also fall below our modulation index cutoff ($m_{\mathrm{I,max}} < 2$). A train of normal giant pulses is clearly visible along the dispersion measure of the Crab, DM = 56.71 pc cm$^{-3}$, and the supergiant pulse is located near $t = 50$ s. This pulse is so bright that it has some of the same characteristics as broadband RFI; namely, events occurring at high dispersion measures. But as we saw previously for J1928+15, applying a $m_{\mathrm{I,max}}$ eliminates many of the spurious events at incorrect DMs because they have low filling factors.

The bottom frame plots $\mathrm{SNR_t}$ vs. $m_I$ (top) and a histogram of $m_{\mathrm{I}}$ (bottom). The events associated with the supergiant pulse are shown in medium gray and fall along the curve extending from $10^4 > \mathrm{SNR_t} > 100$ and $0.7 < m_{\mathrm{I}} < 11$ as well as the clump at $\mathrm{SNR_t} \sim 10$ and $m_{\mathrm{I}} \sim 15$. Note that the largest modulation indices are consistent with the value for spiky RFI ($m_{\mathrm{I,s}} \approx 21$). The correctly dedispersed normal giant pulses are shown in black and lie along the left edge of the main cluster ranging from $100 > \mathrm{SNR_t} > 5$ and $0.5 < m_{\mathrm{I}} < 3$. One exception is the single, larger black point at the low-$m_{\mathrm{I}}$ end of the supergiant pulse curve (medium gray) showing the point associated with correct dedispersion value. Not surprisingly this event has the largest $\mathrm{SNR_t}$ and lowest $m_{\mathrm{I}}$. The events from incorrectly dedispersing the regular giant pulses (light gray) spread to lower $\mathrm{SNR_t}$ and higher $m_{\mathrm{I}}$ than their correctly dedispersed counterparts (black). Finally the noise (light gray) falls between $2 \lesssim m_I \lesssim 7$ and along $\mathrm{SNR_t} \sim 5$ and is consistent with $m_{\mathrm{I,T}} \approx 4.2$.

The measured values of $m_{\mathrm{I}}$ for the Crab pulses are systematically larger than

that predicted by Equation 3.13 for broadband pulses. The brightest regular giant pulses have $SNR_t \approx 100$ and $m_I \approx 0.5$, but the modulation index of a perfectly flat pulse with this $SNR_t$ is $\sim 0.2$. This suggests that Crab giant pulses have significant inherent frequency strucutre and non-zero $m_A$. Most striking is that the modulation index for the supergiant pulse ($m_I \approx 0.7$) is larger than the modulation index for the strongest regular pulses even though its SNR is two orders of magnitude larger. The spectrum of the supergiant pulse shows significant variation across the band due to DISS, as shown in Cordes et al. (2004).

The histogram of $m_I$ shows the total number of events at all DMs (thick line) and the number of events at the DM of the Crab (thin line) as an proxy for the number of giant pulses. The assumption that any event at the DM of the Crab is a giant pulse is simplistic and may include false positives. Because brighter giant pulses are rarer than weaker pulses, Equation 3.13 tells us that the number of pulses in a $m_I$ bin will decrease as $m_I \to 0$. There is therefore a tradeoff between choosing a higher $m_{I,max}$ to allow for the more common, weaker pulses and the false alarm rate.

Figure 3.7 further explores the role of interstellar scintillations on the modulation index. The $SNR_t$ vs. $m_I$ is plotted for two higher frequencies from Cordes et al. (2004). The data are time-frequency snapshots of pulses determined from previous processing at 1475 and 2850 MHz. The snapshots were only dedispersed at the dispersion measure of the Crab pulsar, the time series were thresholded with $SNR_{min} = 3$, and the resulting candidates were run through the cluster algorithm to find the sample with the maximum SNR. Also plotted is the $m_{I,bb}$ curve given by Equation 3.13 with $N_\nu = 118$ at 1475 MHz (top) and $N_\nu = 58$ at 2850 MHz (bottom).

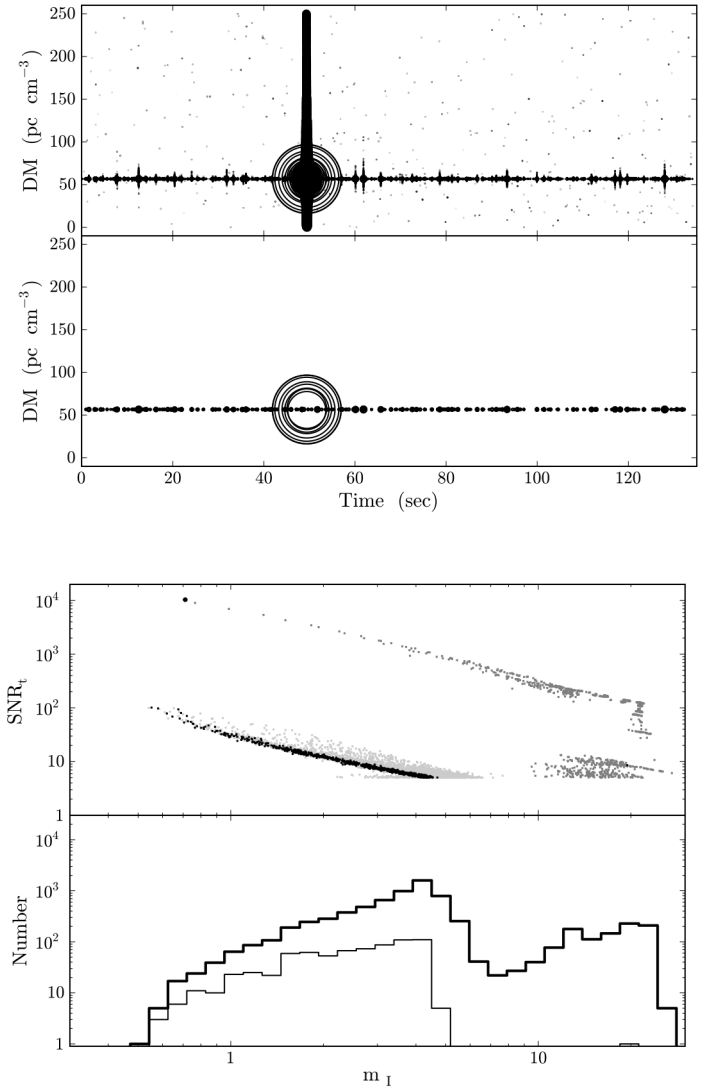


Figure 3.6: Modulation index calculation for the Crab pulsar at 430 MHz. In all frames only events with SNR > 5 are shown. Top Frame: Events above the intensity threshold are plotted against DM and time (top) and with the additional constraint $m_{\mathrm{I,max}} < 2$ (bottom). A stream of Crab giant pulses can be seen in the DM-time plot at DM=56.71 pc cm$^{-3}$. The supergiant pulse discussed in Cordes et al. (2004) occurs around $t = 50$ s. Bottom Frame: The SNR$_t$ vs. $m_I$ is plotted in the top panel. The points associated with the supergiant pulse are plotted in medium gray, points associated with the dispersion measure of the Crab are shown in black, and all other points shown in light gray. A $m_I$ histogram (bottom) plots the total number of events (thick line) and the number of events at DM=56.71 pc cm$^{-3}$ (thin line) as a proxy for the number of pulses detected.

In both panels there are two clusters of points: one due to thresholded noise and one due to the pulses. The thresholded noise lies below $m_{\rm I} = 2$ and $m_{\rm I} = 1.5$ at 1475 and 2850 MHz respectively. The typical values for the noise are consistent with $m_{\rm I,T}$ and lie along the broadband curve as predicted in Section 3.2.3. The modulation indices of the pulses are systematically larger than what is predicted, suggesting the pulses have inherent frequency structure and $m_{\rm A} \neq 0$. The spectra for the strongest pulse at each frequency are shown in Cordes et al. 2004, Figures 7 and 9, and in both cases they exhibit frequency structure caused by DISS.

## 3.4 Extensions of the Method

Our discussion has focused the detection of radio bursts using incoherent dedispersion, but the technique has more general applicability. It can also be applied to other classes of signals (periodic objects like pulsars, spectral lines) and other data formats (coherent dedispersion, image cubes from interferometers).

### 3.4.1 Periodic Signals

Using the modulation index to characterize the frequency structure of a single pulse can be extended to periodic signals. The procedure is best illustrated by considering a dedispersed time series $\bar{I}(t, DM)$ that has been folded at some period $P$. A pulse in this folded time series is the sum of $N_{\rm p}$ pulses

$$\bar{I}_p(t_{\rm o}, DM) = \frac{1}{N_{\rm p}} \sum_{\rm t_p} \bar{I}(t, DM) \tag{3.21}$$

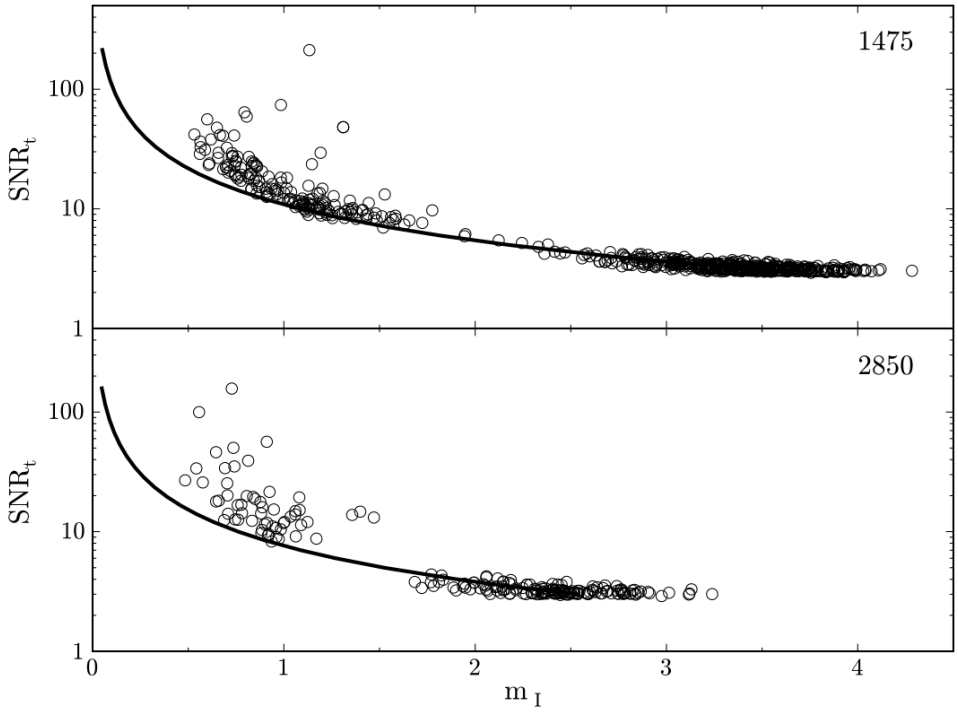


Figure 3.7: $SNR_t$ vs $m_I$ of Crab giant pulses at 1475 MHz (top) and 2580 MHz (bottom) from Cordes et al. (2004). The lines represent $m_{I,bb}$ (Equation 3.13). The modulation index that divides the pulses from thresholded noise is $m_I \approx 2$ for 1475 MHz and $m_I \approx 1.5$ for 2850 MHz. The modulation indices of the thresholded noise are consistent with those of broadband pulses. The modulation indices of the pulses are systematically larger than what is predicted by Equation 3.13, suggesting the spectra have inherent structure caused by diffractive interstellar scintillations.

where $t_{\rm p} = t_{\rm o} + nP$, $t_{\rm o}$ is the time sample of the candidate pulse in the folded time series, and $n = 0 \ldots N_{\rm p} - 1$. The modulation index is then calculated for samples where $\bar{I}_p(t, DM)$ is above the SNR threshold. As discussed in Section 3.2.1 for smoothing in time, calculating the second moment requires that spectra are first summed in time before averaging over frequency, making it obligatory that one returns to the time-frequency data

$$\overline{I_p^2}(t_{\rm o}, DM) = \frac{1}{N_\nu} \sum_{\nu} \left[ \frac{1}{N_{\rm p}} \sum_{t_{\rm p}} I(t, \nu, DM) \right]^2 . \tag{3.22}$$

One can also calculate the second moment for the Fourier analysis method of finding pulses signals but that is beyond of the scope of this paper.

The additional computational cost required to calculate the modulation index can be estimated in the same manner as for the single pulse detections described in Section 3.3.2. But as a pulsar is a repetitive source, the time span of the data that would need to be reprocessed in a second pass would be larger than for a single pulse. In the extreme case where $N_{\rm s} = N_{\rm t}$, the processing load is comparable when the number of candidates is of the order of the number of trial DMs. It is therefore prudent to reduce the number of candidates by other means first before calculating the modulation index of candidate pulsar.

### 3.4.2 Coherent Dedispersion

Coherent dedispersion operates on the basedband voltage directly rather than on intensity-like quantities at the output of a spectrometer, which our previous discussion has centered on. In a survey, a set of trial values for DM would be used, as with post-detection dedispersion described earlier. With the correct value of DM, coherent dedispersion restores the phase of the signal to what it

was prior to alteration by the ISM (so long as multipath scattering is not important). The procedure deconvolves a phase function from voltage data $\epsilon(t)$ by applying a complex phase filter (e.g., Hankins & Rickett, 1975) to produce the dedispersed voltage $\epsilon_d(t)$. The detected signal $I_d(t) = |\epsilon_d(t)|^2$ would then be analyzed with a SNR threshold to identify candidate bursts.

If the true signal is an unresolved pulse with width $W \ll B^{-1}$ where, as above, $B$ is the total bandwidth, the dedispersed output will have a width $W_t \approx B^{-1}$ only if the signal is unmodulated across the band. This requires that any scintillation modulation have a characteristic bandwidth $\gg B$. In this instance, no new information is gained by analyzing the data in the combined frequency-time plane.

For signals that have broader time extent, however, the analysis presented in the paper for post-detection dedispersion still applies. The dynamic spectrum would be calculated for each burst identified in $I_d(t)$ from short time blocks of length $\Delta t_s$ using the discrete Fourier transform (DFT) $\tilde{\epsilon}(t)$ of the baseband signal. Using an $N_\nu-$point DFT, the uncertainty relation implies $\Delta t_s B/N_\nu = 1$. For burst widths $W \gtrsim \Delta t_s$, the series of spectra that span the burst can be used to calculate the modulation index across frequency, which is then used to classify signals as before.

### 3.4.3 Application to Images Obtained On a Frequency-Time Grid

The modulation index as discussed so far is calculated for a set of intensity measurements sampled in time and frequency at a particular value of the dispersion measure, $I(t, \nu, DM)$. For signals that are not inherently narrow in time (i.e., not pulsed or transient in nature), no dedispersion is required. Instead, we simply calculate the modulation index $m_{\rm I}$ as the normalized variance of the intensity across frequency at a particular time,

$$m_I^2 = \frac{\overline{I^2}(t) - \overline{I}(t)^2}{\overline{I}(t)^2}. \tag{3.23}$$

The modulation index $m_{\rm I}$ and fractional correlation bandwidth FCB calculated in this manner can be used, for example, to characterize radio observations of a spectral line or maser source (which show coherent frequency structure) and discriminate them from radio frequency interference with a more random frequency structure. We discuss the application of a time-domain modulation index to these source types in Section 3.5.

The method can be further generalized to apply to interferometric imaging observations of continuum sources that have been acquired in multi-channel modes. While past imaging observations with the Very Large Array (VLA), for example, have only employed a small number of channels (e.g., $\lesssim 32$) or even just one channel in continuum mode, observations with the Expanded Very Large Array (EVLA) or the future SKA pathfinders such as Austrailan Square Kilometre Array (ASKAP; Johnston et al. 2009) and MeerKAT (Booth et al., 2009) will typically employ many more channels. At the EVLA, the wide bandwidth available for continuum imaging (e.g., 1–2 GHz or 2–4 GHz) requires the use of

many spectral channels to avoid chromatic aberration ("bandwidth smearing"). If we were to require a maximum tolerable peak response reduction of 20% at A-array, 1–2 GHz, that would necessitate a channel bandwidth $\Delta\nu$ such that

$$\frac{\Delta\nu}{1.5\ \mathrm{GHz}}\frac{\theta}{\theta_0} < 1,$$

where $\theta/\theta_0$ is the source offset from the phase tracking center in units of the synthesized beam.[3] To maintain sensitivity at that level over just half the primary beam field of view, $\theta/\theta_0 \sim 0.5 \times d/D$ where $d$ and $D$ are the dish diameter and array size respectively. For the EVLA in A-array where $d = 25$ m and $D \approx 36$ km, this constraint requires as many as 720 channels for continuum imaging.

One of the challenges faced by automated source extraction pipelines working on these data is to discriminate between low significance detections of compact sources and random intersections of the point spread function sidelobes caused by strong sources in the field of view. While the exact shape of the sidelobes depends on the details of the array and the observation, they will scale as $I(\theta, \nu) \propto \nu^{-1}$ at an angular distance $\theta$ from a bright source. Therefore, an intersection of sidelobes from different sources will show smooth structure in amplitude as a function of channel frequency, while a true compact source will not. Calculating the modulation index and FCB will allow us to exploit the extra information in the frequency channelization of synthesis images and add an extra discriminant that is easily implementable in automatic pipelines.

[3] See, e.g., the EVLA Observational Status Summary, `http://evlaguides.nrao.edu/`

### 3.4.4 Realtime RFI excision

The modulation index could be useful for systems performing RFI excision on the fly. Interest in realtime incoherent dedispersion transient searches is growing thanks to the large speed-ups achieved by graphics processing units (GPUs) (e.g. Magro et al., 2011). As more of the signal processing moves into hardware, realtime RFI rejection is also moving into instruments. For example Deller (2010) have incorporated the calculation of kurtosis into the F-stage of the DiFX software correlator, and the modulation index could be calculated in an identical manner.

In the time-frequency domain the modulation index would be most useful for identifying broadband, impulsive RFI. A sample that had an anonymously low modulation index could be flagged or blanked by the hardware. The modulation index alone should not be used to identify narrowband RFI, as a dispersed pulse is narrowband before the dispersion is correctly accounted for.

## 3.5 Discussion

In this paper we have focused our discussion on searching for radio bursts from or similar to those from pulsars and used the modulation index, along with SNR, as a tool for RFI excision. More generally the modulation index can be one among several statistics used to *characterize and classify* signals rather than to filter candidates. Furthermore there may be as-yet undiscovered source classes having different time-frequency signatures.

Low mass stars (M-dwarfs and brown dwarfs) are frequent emitters of radio

bursts (Jackson et al., 1989; Berger, 2002). Bastian et al. (1990) have shown that the dynamic spectra of dMe flare stars exhibit a rich structure in both time and frequency. The spectra of the flares they observed had both broadband and narrowband components, as well as large spectral indices ($\alpha \approx 10$), suggesting various plasma phenomena at work. A large-scale survey of flare stars could use the modulation index as one of several parameters used to automatically classify flare types. Furthermore, different plasma processes result in different levels of polarization, so calculating the modulation index of each polarization separately could further aid in classifying flares and recognizing RFI.

The study of radio variability in the solar system is a natural extension to flare stars and is another area where the modulation index could be a tool for classifying different radio bursts. Planetary auroral radio emission (ARE) has been detected for all of the magnetic planets in our solar system (Zarka, 1998a), and dynamic spectra of ARE show complex substructure in both time and frequency. A notable class of fast radio pulses is Jupiter's "S-bursts". These bursts have durations of milliseconds and are the brightest of Jupiter's ARE emission (Zarka, 1998a). Radio discharges from lightning on the gas giant planets have durations on the order of 10 to 100 ms and fluxes easily detectable at Earth (Zarka et al., 2004). In fact Ruf et al. (2009) have used multi-moment techniques (total intensity and kurtosis) to look for lightning on Mars during a dust storm. By extension it is expected that extrasolar planets would have radio emission that would be variable and include bright bursts (Farrell et al., 1999; Grießmeier et al., 2007). A study of the modulation indices of flares from Jupiter could help identify similar flares from exo-Jupiters.

Extragalactic sources of fast radio transients might include merging neutron

star–neutron star or neutron star–black hole binaries (Hansen & Lyutikov, 2001). Short gamma-ray bursts (SGRB) are thought to arise from the merger of two neutron stars, and Pshirkov & Postnov (2010) suggest that GRB monitors could alert low frequency radio observatories (e.g., LOFAR) to a possible radio transient.

Our focus has been on signals that are broad in frequency but narrow in time. The reciprocal problem of spectral lines that are steady in time can be handled in the same way but with the roles of frequency and time reversed. As an example, surveys for maser lines will most likely require a line amplitude that is relatively steady over time scales of days or less. Interstellar scintillation may induce time variations in some maser sources if they are compact enough. SETI (search for extraterrestrial intelligence) often postulates narrowband beacon signals that are constant in time. Interstellar scintillation will certainly induce time variations owing to the compact nature of any relevant transmitters (Cordes & Lazio, 1991) but with a modulation index $\sim 1$. The methods outlined here apply directly to these problems.

## 3.6 Conclusion

We have discussed how defining detection statistics based on higher order moments can improve the success of source detection pipelines and focused on the spectral modulation index. By calculating the second moment, we are able to classify signals based not only on their strength but also on their fractional frequency variation. By applying prior information about our target sources, i.e., that they are broadband, the modulation index easily distinguishes between

broadband and narrowband signals and allows us to filter a large fraction false positives due to narrowband RFI.

These detection statistics (e.g., SNR, $m_{\mathrm{I}}$) are crucial to source extraction pipelines because they can be calculated automatically. As new observatories generate more and more data, it is critical that source extraction occurs reliably with minimal human intervention. Although we limit ourselves to two statistics in this paper, a pipeline could make use of many higher-order statistics (e.g. kurtosis, Nita et al., 2007) along with a weighted voting scheme to classify signals in more sophisticated and nuanced ways. For example the modulation index could become one parameter used by detection pipeline based on an artificial neural network (Eatough et al., 2010).

We thank the reviewer for his or her useful comments. This work was supported by NSF grant AST - 1008213. L.G.S. and J.S. were supported by the NASA/New York Space Grant Consortium under grant NNX10AI94H. L.G.S. also received support from the National Astronomy and Ionosphere Center (NAIC). The Arecibo Observatory is operated by SRI International under a cooperative agreement with the National Science Foundation (AST-1100968), and in alliance with Ana G. Méndez-Universidad Metropolitana, and the Universities Space Research Association.

CHAPTER 4

# ARECIBO OBSERVATIONS OF A 13-MIN COMPACT, ECLIPSING WHITE DWARF BINARY

The white dwarf binary system J0651 was discovered in a survey for extremely low mass white dwarfs (Brown et al., 2010; Kilic et al., 2011; Brown et al., 2012). Its orbital period is only 12.75 min and has a gravitational merger time scale of 0.9 Myr (Brown et al., 2011). The system also eclipses, allowing for the precise determinations of the masses. If the white dwarfs are magnetic, there is a possibility for magnetic interactions to produce radio emission. One possible model for such emission is a uni-polar inductor Willes & Wu (2004). We present observations of J0651 conducted at the Arecibo Observatory at three frequencies: 327 MHz, 1.5 GHz, and 4.6 GHz. Between 1.5 and two hours were spent observing J0651 at each frequency, and the observing mode was specifically designed to look for variable emission. In particular we focused on short duration, burst-like emission and emission that varied with the orbital period of the binary. No emission was detected, but our efforts were hampered by low frequency noise. We give lower limits and discuss their implications.[1]

## 4.1 Introduction

Surveys for extremely low mass (ELM) white dwarfs (WD) have found close to 20 double degenerate binary systems (Brown et al., 2010; Kilic et al., 2011; Brown et al., 2012). The system with the shortest orbital period discovered by

[1] L.G. Spitler, J. Cordes, S. Chatterjee (Cornell University), W. Brown (SAO), M. Kilic (University of Oklahoma)

the ELM survey to date is SDSS J065133.33+284423.3 (hereafter J0651) (Brown et al., 2011) with an orbital period of 12.75 minutes. The primary is a helium core WD with a mass and radius of $M = 0.25M_{\odot}$ and $R = 0.0353 \pm 0.004R_{\odot}$, and the secondary is a likely carbon-oxygen WD with a mass and radius of $M = 0.55M_{\odot}$ and $R = 0.0132 \pm 0.0003R_{\odot}$. The system is eclipsing, and a period of $P = 765.2062 \pm 0.003$ s is determined from the light curve. J0651 lies at a distance of $1.0 \pm 0.1$ kpc from the Sun, and there is no evidence for mass transfer, making J0651 a detached binary. With a predicted gravitational merger rate of 0.9 Myr, it is one of the strongest emitters of gravitational waves currently known. Recently Hermes et al. (2012) observed a decrease in the orbital period matching the predicted decrease from the emission of gravitational radiation.

The presence of radio emission is common in astrophysics when magnetic fields are present. In many cases the radiation is coherent. All of the solar system's planets with magnetic fields emit radio emission (Zarka, 1998b), including short-duration bursts such as the Jovian S-bursts that have high brightness temperatures and millisecond durations. Low mass stars (M-dwarfs and brown dwarfs) are frequent emitters of radio bursts (Jackson et al., 1989; Berger, 2002). These bursts are believed to be electron cyclotron maser (ECM) emission (Hallinan et al., 2006). Pulsars are the classic example of radio emission from compact objects arising from complicated emission mechanisms involving strong magnetic fields.

J0651 is likely similar to the two short-period X-ray sources, RX J1914+24 (Cropper et al., 1998; Ramsay et al., 2000) and RX J0806+15 (Israel et al., 1999; Ramsay et al., 2002), that exhibit periodic emission at P = 9.5 min and P = 5.4 min respectively at both X-ray and optical. Both orbital periods are too short

to allow a WD and normal star, so these systems are likely double degenerate. The lack of periodic emission at any other periods and the 50% duty cycle of the X-ray emission suggests the periods observed in the X-ray and optical should be associated with the orbital period and not an accretion disk.

Wu et al. (2002) proposed an unipolar inductor model to explain the X-ray and optical emission. In a unipolar inductor an emf is induced in a conductive object moving through a magnetic field, generating a current along the flux tubes to the magnetic object. The most well-known example is the Jupiter-Io system (Piddington & Drake, 1968; Goldreich & Lynden-Bell, 1969). In a binary WD system, one WD must be magnetic and the other non-magnetic. According to the unipolar inductor model, the X-ray and optical emission in RX J1914+24 and RX J0806+15 arises from resistive heating in the atmosphere of the magnetic white dwarf at the footprints of the magnetic flux tube (Wu et al., 2002; Dall'Osso et al., 2007). Similar footprints are seen in the ultraviolet in Jupiter's atmosphere.

A unipolar inductor can also generate electron cyclotron maser emission (ECM) detectable at radio wavelengths. ECM emission has high brightness temperatures ($T_b \gg 10^8$) and is 100% circularly polarized, providing a clear observational signature. Again looking to the Jupiter-Io system, ECM is responsible for the Jovian DAM emission at $\sim$ 10 MHz (Zarka, 1998b). Willes & Wu (2004) model ECM in a system containing a magnetic and non-magnetic WD with properties similar to J0651, and find that over a range of model parameters, radio emission will be generated with flux densities ranging from 10 $\mu$Jy to 0.1 Jy at 5 GHz, although Lai (2012) suggests caution with such estimates.

Ramsay et al. (2007) searched for ECM emission from RX J1914+24 and RX

J0806+15 using the VLA at 5 GHz. RX J1914+24 was not detected with a 3-$\sigma$ upper limit of 99 $\mu$Jy. RX J0806+15 was observed on two sessions separated by 15 months. In the first session a point source coincident with the optical position was detected at 5.8-$\sigma$ with a flux density of 99 $\mu$Jy and $\sim$ 50% circular polarization (see Ramsay et al. (2007) for caveats on this measurement). The second observation of RX J0806+15 resulted in a non-detection and a 3-$\sigma$ upper-limit of 36 $\mu$Jy. The lack of redetection suggests these sources are intrinsically variable over long time scales and target sources of slow transient surveys.

With this motivation in mind, we observed J0651 at multiple radio frequencies with the Arecibo Observatory to look for variable emission. In Section 4.2 we describe the observations. In Section 4.3 we discuss our data analysis techniques and describe the results. The implications of our results and final thoughts are given in Section 4.4 and Section 4.5 respectively.

## 4.2 Observations

Observations of the binary white dwarf J0651 were undertaken at the Arecibo Observatory to search for periodic or burst-like emission. Data were taken with the 327 MHz receiver, the L-wide receiver, and the C-band receiver on 9 March 2012, 10 March 2012, and 11 March 2012 respecitively using the PDEV/Mock[2] spectrometers (hereafter simply Mock spectrometers). The spectrometers recorded 16-bit spectra every 1 ms with full polarization information. With the 327 MHz receiver we used a single Mock spectrometer with a center frequency of 326.9 MHz, 256 frequency channels, and a bandwidth of 57.3

[2] http://www.naic.edu/~phil/hardware/pdev/usersGuide.pdf

MHz. At the higher two frequency bands, the available receiver bandwidth was larger than the bandwidth of a single Mock spectrometer, so multiple spectrometers were used to span the band. Each spectrometer provided a subband with 172 MHz of bandwidth and 512 frequency channels, and the center frequencies of the subbands were chosen to give an overlap of approximately 20 MHz to avoid the roll-off at the edges of each Mock subband. The L-wide receiver required four Mock spectrometers spanning a total bandwidth of 623.95 MHz centered on 1440.3 MHz. The C-band receiver required seven Mock spectrometers spanning a total bandwidth of 1075.5 MHz centered on 4616.1 MHz.

At the start of each observing session, the pulsar B0611+22 was observed for two minutes to verify data integrity. The flux calibrator J0642+214 was observed for five, 30-second pointings while switching the calibration diode on and off with a 50% duty cycle and two second period. One pointing was directly on the source, and the other four were offset in right ascension and declination by three beam widths. The rest of each observing session was spent tracking the white dwarf binary with between 1.75 and 2 hours on source. The calibration diode was turned on for 5 s every 60 minutes.

## 4.3 Data Analysis

First, the data spanning multiple Mock spectrometers are merged by discarding the overlapping frequency channels at the edge of each subband and concatenating the remaining samples. In the case of the 1.5 GHz observations, the lowest Mock subband was excluded in its entirety because it was overwhelmed with radio frequency interference (RFI). The resulting data parameters for each

observation frequency after merging are given in Table 4.1.

We inspect the data by generating dynamic spectra of the time-frequency data, as well as time series formed by averaging over frequency. Before the data are inspected, several processing steps must be taken. We correct for the known increase in system temperature at zenith angles greater than about 15 degrees by scaling the data according to calibration curves [3]. The bandpass shape is flattened by defining a characteristic bandpass function from the data themselves. First, a set of spectra was defined by calculating the median spectrum of 26-second long blocks of data. After removing outliers from this set, the remaining spectra were averaged to give the characteristic bandpass function. The data were flattened by dividing by this function. Finally, several RFI rejection algorithms were applied as described in detail below.

### 4.3.1 RFI rejection

The RFI rejection algorithms operate on 1-s long chunks of frequency-time data and target three classes of interference: persistent or burst-like narrowband RFI, and bursty-like, broadband RFI . Persistent narrowband RFI was flagged after inspecting the data by hand for channels that are high over a large fraction of the observation. Persistent, narrowband RFI occurred at all three frequencies. The data at 4.5 GHz were relatively free of burst-like RFI, and RFI excision was largely done manually.

The data collected at 327 MHz and 1.5 GHz required more thorough RFI cleaning. Short-duration, broadband RFI was common in the 327 MHz data

[3] http://www.naic.edu/~phil/gaincurves/gaincurves.html

and occurred occasionally at 1.5 GHz. It was identified using the spectral modulation index technique (Spitler et al., 2012), which provides a measure of how narrow or broad band a signal is. Spectra with modulation indices that suggest the presence of a bright, broadband signal where flagged as "bad" and a padding of ten "bad" samples was added on either side. These bad spectra were replaced with the median of the "good" samples in the 1-second block. Figure 4.1 shows dynamic spectra along with a time series and average bandpass for a 10 s sample of the data at full time resolution. The top panel shows the data after correcting for system temperature variations and flattening the bandpass but with no RFI rejection. Both persistent narrowband RFI is visible as well as a bright, short duration burst of RFI. The bottom panel illustrates how our RFI rejection algorithms excise the RFI.

Short-duration, narrowband RFI was common at both 327 MHz and 1.5 GHz. It is identified by finding frequency channels in the 1-s chunks that are outliers. Outliers are defined as any channel whose SNR is 50% larger than the SNR expected for one false positive. These bad channels are replaced by the median of the entire 1-sec block. The 1.5 GHz data are also plagued by airport radar with a six second period. These were identified by applying a 2D sigma-clipping routine. Finally, if the fraction of data flagged within a one-second block is greater than 75%, the entire block is replaced with the median of that block. Figure 4.2 shows the three panel plot described above for 20 s of data at 1.4 GHz without RFI rejection (top) and with RFI rejection (bottom). Clearly persistent, narrowband RFI is a major problem at 1.4 GHz. A bright radar pulse can also be seen near t = 3141 s.

After applying all of the RFI algorithms to the frequency-time data, a time

| Receiver | $\nu_{ctr}$ | BW | $N_\nu$ | $T_{tot}$ | $N_{orb}$ | MJD |
|---|---|---|---|---|---|---|
| 327 MHz | 326.9 MHz | 50 MHz | 224 | 6580 s | 8.6 | 55995.956 |
| L-wide | 1507.4 MHz | 444.9 MHz | 1324 | 6953 s | 9.1 | 55996.952 |
| C-band | 4599.0 MHz | 997.9 MHz | 2970 | 7085 s | 9.3 | 55997.949 |
| C-band$^\dagger$ | 4599.0 MHz | 997.9 MHz | 2970 | 4600 s | 6.0 | 55998.449 |

Table 4.1: Data parameters. The table columns are as follows: $\nu_{ctr}$ is the center frequency of the data after merging the Mock subbands, BW is the total bandwidth, $N_\nu$ is the number of frequency channels, $T_{tot}$ is the total observing time on J0651+28, $N_{orb}$ is the number of orbits of J0651 in the timespan, and MJD is the modified Julian date at the start of the observation of J0651+28. C-band$^\dagger$ gives the parameters after excluding the time spans most affected by the problems with the C-band receiver.

series is calculated by averaging over frequency. Ranges of known bad data (i.e. calibration diodes, extreme RFI) are masked. A polynomial is subtracted from the time series to remove long-time scale drifts in the system temperature. For the 4.6 GHz data, a first-order polynomial was used, and for the other two frequencies, a third-order polynomial was used.

After the observations were conducted, we learned of technical issues with the C band receiver that result in large swings in system temperature (Phil Perillat, private communications). Furthermore, the increase in system temperature is accompanied with a steepening of the spectrum toward higher frequencies. Figure 4.3 shows these variations in system temperature The worst variations occurred in the first 1800 s of the observation and in the last 685 s, and we have excluded these ranges from our analysis. The resulting time series is 4600 s long, corresponding to 6 orbits of the system. Two other weaker, short-duration system temperature increases associated with a steepening in the bandpass were seen during this 4600 s time span. One could be efficiently removed by subtracting off a polynomial, and the second was masked.

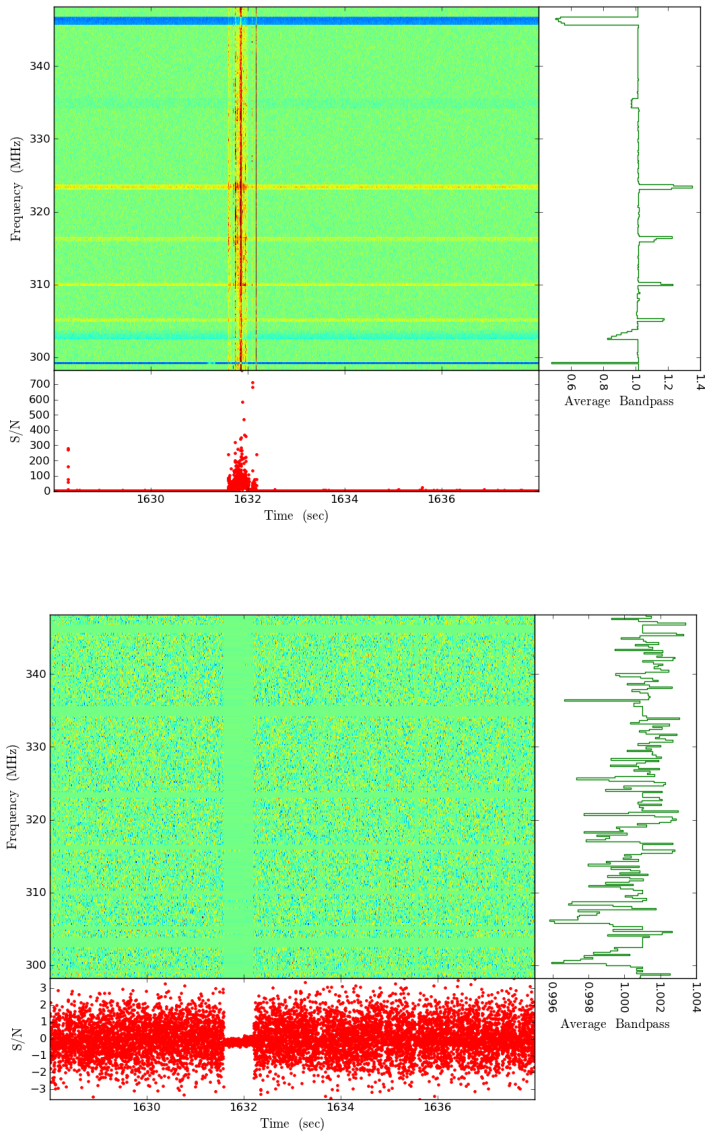


Figure 4.1: Dynamic spectra, time series (bottom), and average bandpass (right) at 327 MHz for 10 seconds of data at full time resolution. The top panel shows the raw data with no RFI excision algorithms applied. The lower panel shows the cleaned data after applying RFI excision algorithms.

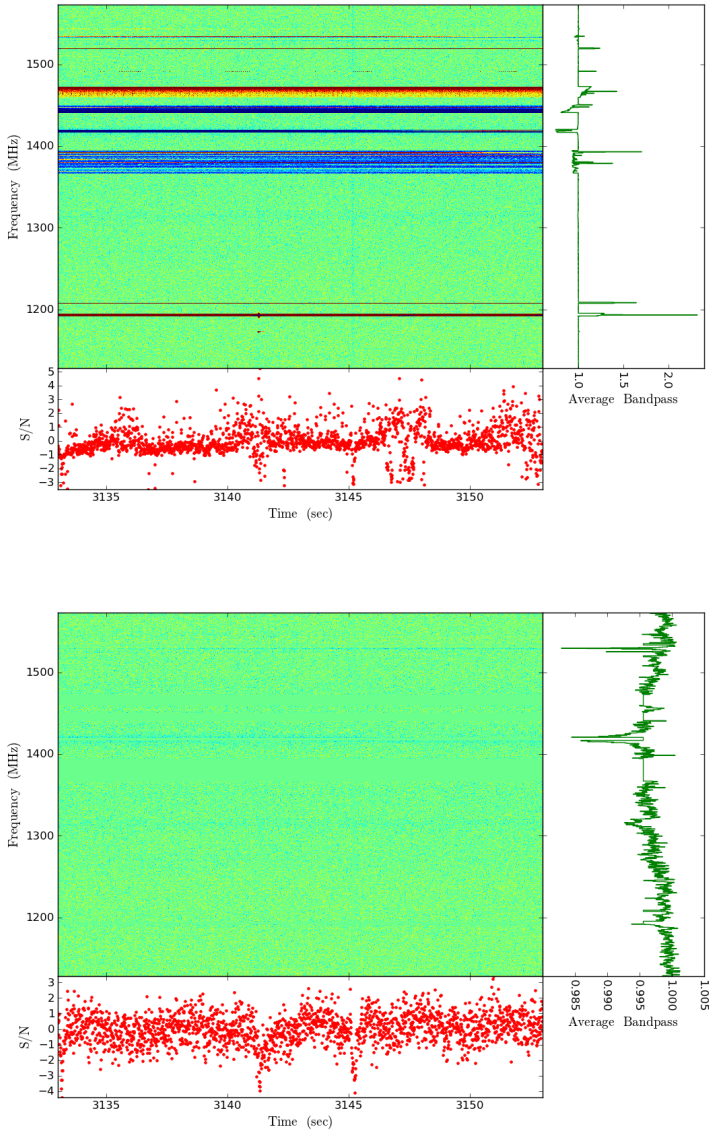


Figure 4.2: Dynamic spectra, time series (bottom), and average bandpass (right) at 1.4 GHz for 20 seconds of data with 10 ms time resolution. The top panel shows the raw data with no RFI excision algorithms applied. The lower panel shows the cleaned data after applying RFI excision algorithms.

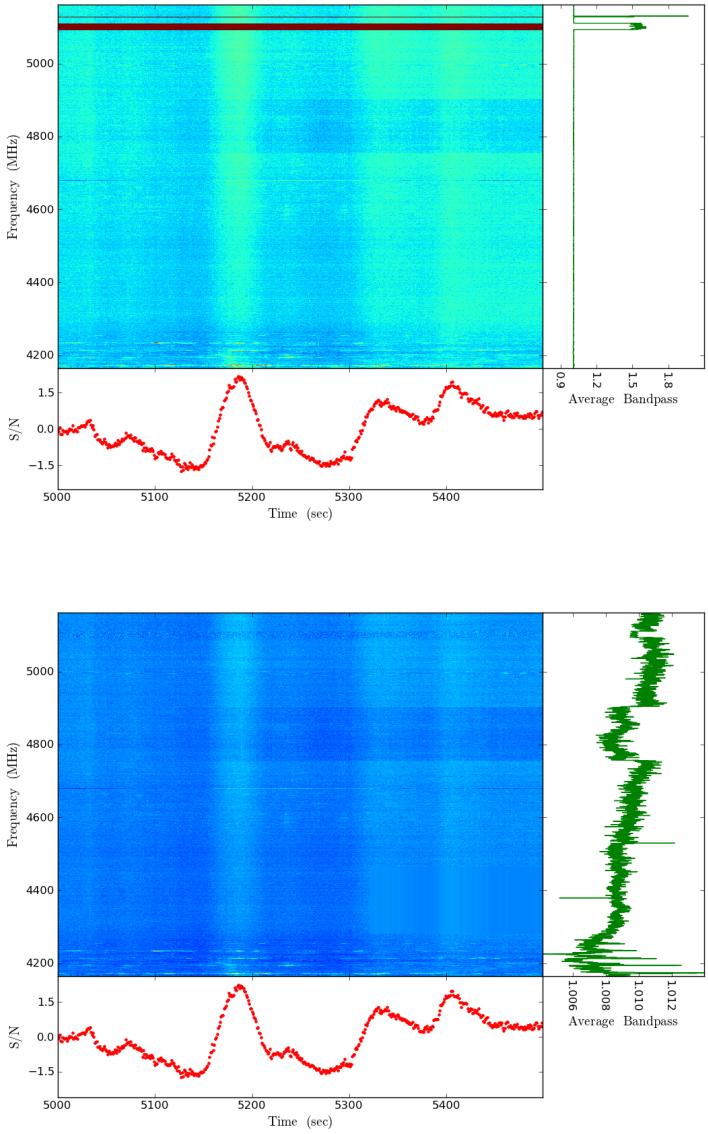


Figure 4.3: Dynamic spectra, time series (bottom), and average bandpass (right) at 4.6 GHz for 500 seconds of data with 1 s time resolution. The top panel shows the raw data with no RFI excision algorithms applied. The lower panel shows the cleaned data after applying RFI excision algorithms.

### 4.3.2 Burst analysis

We looked for bursts in the time series using a matched filtering algorithm (Cordes & McLaughlin, 2003). The time series at each frequency was convolved with a template bank of 15 boxcar filters with effective widths of 1 ms - 1 s. The exact widths of the filters were 1, 2, 3, 4, 6, 9, 14, 20, 30, 45, 70, 100, 150, 300, 500, 1000 ms. A list of candidate pulses was created by applying a threshold of $\mathrm{SNR}_{\mathrm{min}} > 5$ to each of the convolved time series. We then looked at the candidate pulses in the time-frequency plane to test whether they were caused by radiometer noise or RFI.

We can place a $5\sigma$ upper limit on the presence of short-duration bursts using the following formula:

$$S_{\mathrm{max,W_t}} = S_{\mathrm{max,1ms}} \sqrt{\frac{1\mathrm{ms}}{W_t}} \left(\frac{m}{5}\right), \tag{4.1}$$

where $W_t$ is the temporal width of a burst, $m$ is the standard deviation multiplier, $S_{\mathrm{max,W_t}}$ is the upper limit to a burst width $W_t$, and $S_{\mathrm{max,1ms}}$ is the upper limit to a burst with a width of 1 ms, the values of which are given for each frequency band in Table 4.2. The $1\sigma$ flux density error was determined from the $1\sigma$ uncertainty in the flux density conversion factor.

We also looked the possibility of bursts by looking at dynamic spectra of the data with a range of different time resolutions (10 ms - 10 s). No candidate bursts were found.

| Band | $S_{\mathrm{max,1ms}}$ (mJy) | $S_{\mathrm{max,per}}$ (mJy) | $\sigma_{\mathrm{S}}$ ($\mu$ Jy) |
|---|---|---|---|
| 327 MHz | 80.0 | 4.3 | 300 |
| 1.5 GHz | 7.8 | 2.2 | 72 |
| 4.6 GHz | 4.0 | 2.5 | 23 |

Table 4.2: Flux density upper limits for variable emission.

### 4.3.3 Periodicity analysis

The most robust detection of emission from the binary WD system would be a signal modulated at the binary orbital period. We used two different techniques to look for periodic emission: Fourier analysis and folding the time series at a series of trial periods. Figure 4.4 shows the Fourier transforms of the time series each of the three frequency bands (from top to bottom: 327 MHz, 1.4 GHz, and 4.6 GHz). The top panel in each figure shows the entire spectrum, and the bottom panel expands the low frequency part of each spectrum. The dashed vertical line marks the orbital frequency of the binary system. At higher frequencies only RFI is visible (e.g. 60 Hz and its harmonics). The bottom panels show significant power at low-frequencies, which suggests variations in the baseline have not been properly accounted for. There is no obvious spike at the system's orbital frequency. We also applied a high pass filter to the time series with a frequency cutoff of half the orbital frequency. While this reduced the level of the lower frequency power, it revealed no obvious signal at the orbital frequency.

We also investigated the presence of variable emission by folding the time series at the orbital period and inspecting the time series for each individual orbit. These orbit stacked plots are shown in Figure 4.5. The top panel in each figure is the folded, binned time series. Each subsequent panel shows the data

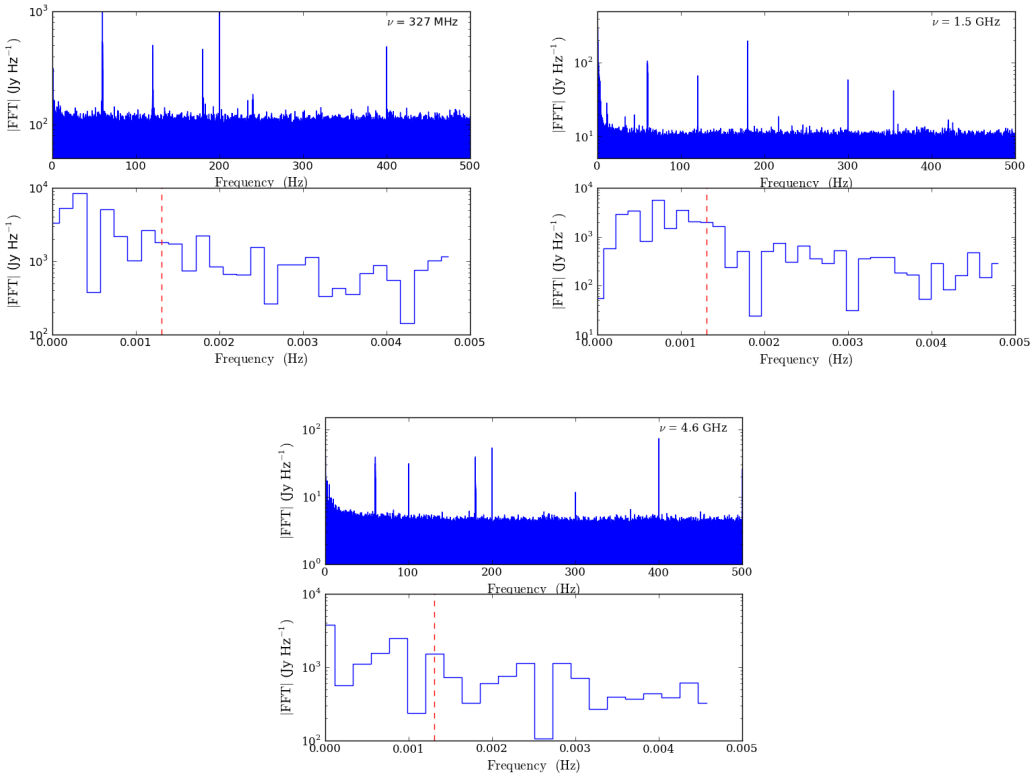


Figure 4.4: Fourier transforms of the time series at each of the three observing frequencies. From top to bottom is the 327 MHz, 1.5 GHz, and 4.6 GHz transforms. For each figure the top panel shows the transform of the entire frequency range of the time series spectrum. The bottom panels zoom in on low frequencies. The orbital frequency of the binary system is shown by the vertical red dashed line. Significant power is seen at all three frequencies, but no obvious increase exists at the system's orbital frequency.

comprising each orbit. While variable emission is clearly seen, there is no obvious consistency from orbit to orbit in any of the orbital phase bins. The gaps in the time series correspond to data that were flagged for a various reasons (e.g. extreme RFI, calibration diodes).

We further investigated deviations from Gaussian white noise by folding the time series at the orbital period, binning to 1s phase bins, and calculating the re-

duced $\chi^2$. The reduced $\chi^2$ for each band was 26.4 at 327 MHz, 687.7 at 1.5 GHz, and 2644.5 at 4.6 GHz, which suggests large deviations from Gaussian noise. But we already know from the Fourier analysis that there is significant low frequency power, so we folded each time series over a range of trial periods from 100 - 1000 s in 10 second intervals, binned to 1 s phase bins, and calculated the reduced $\chi^2$ for each trial period. The reduced $\chi^2$ was not significantly larger when folded at the orbital period than at the other trial periods, so we conclude that the large reduced $\chi^2$ at each frequency band arrises from instrumental variations, not astrophysical emission.

## 4.4 Discussion

There are a number of possible explanations for why we did not detect radio emission, which can roughly be classified as either observational or intrinsic. Our observations were only sensitive to variable emission, not persistent emission. If the source is a steady radio emitter, we would not have been able to recognize it. It is also possible that J0651 is variable, but at a level below our sensitivity, or that it is variable on time scales longer than our observations (such as RX J0806+15). If any emission is caused by the electron cyclotron maser process, the peak emission will depend on the source's magnetic field, and we may not have observed at the optimal frequency band. The magnetic field strengths associated with emission peaking in our frequency bands are 120 G, 540 G, and 1.6 kG at 327 MHz, 1.5 GHz, and 4.6 GHz respectively. The weakest measured white dwarf magnetic field is $\sim$ 2 kG (Aznar Cuadrado et al., 2004), and several authors have found the incidence of DA white dwarfs with magnetic fields of the order $\sim$ 10 - 100 kG to be about 10% (e.g. see Landstreet et al., 2012). These

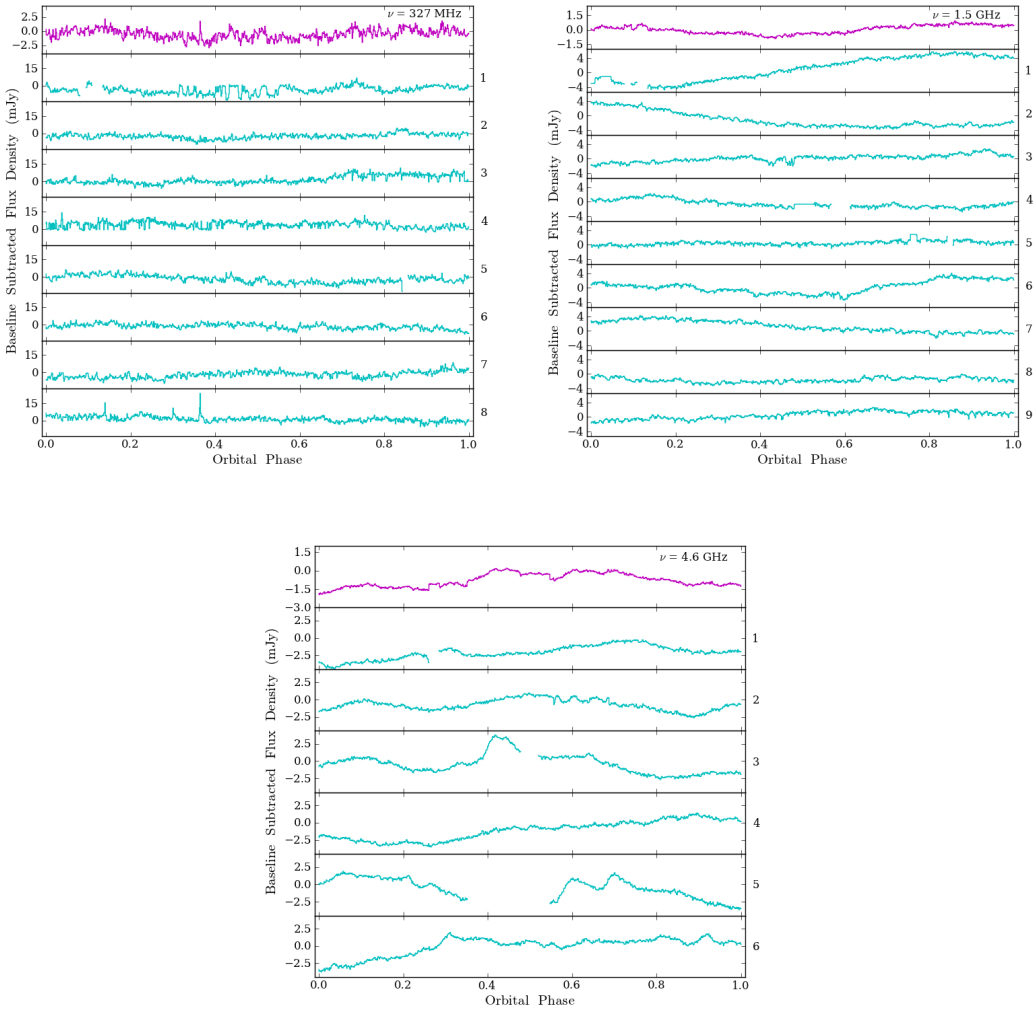


Figure 4.5: Folded time series of binary white dwarf system with each orbit stacked at each of the three frequencies. The three figures from top to bottom are 327 MHz, 1.5 GHz, and 4.6 GHz. The top panel in each figure shows the summed time series binned to 1s phase bins. The other panels from top to bottom show the time series from each of the orbits that comprise the folded time series. The gaps in the orbit time series are regions where the data were flagged due to extreme RFI or to mask the calibration diode.

results suggest that it may be more fruitful to observe at higher radio frequencies to probe higher magnetic field strengths. Whether white dwarfs exist with magnetic fields weaker than a $\sim$ 1 kG exist is an open question, as measuring weaker magnetic fields is technically challenging.

Figure 4.6 summarizes these results. The frequencies bands observed in these observations are plotted against the variability time scales studied in our analysis. On the right is the magnetic field corresponding to the emission frequency shown on the left assuming cyclotron emission. The vertical line denotes the frequency of the binary white dwarf orbit. The short and long time scale limits are set by the sample time of the data and the total observation duration respectively. This figure clearly illustrates that we are sampling a wide range of time scales (almost seven orders of magnitude), while a fairly limited range of magnetic fields is probed (about one order of magnitude). This again argues for expanding the observing bands to higher radio frequencies.

It is also possible that J0651 may not behave like a unipolar inductor. A unipolar inductor requires relative motion between magnetic and non-magnetic white dwarfs. If the source were tidally locked, no potential difference would be set up. However, Fuller & Lai (2012) have shown that dynamical tides can prevent a compact white dwarf binary from becoming tidally locked before merging, so the white dwarfs in J0651 may be rotating asynchronously. The unipolar inductor model also requires one magnetic and one non-magnetic white dwarf. If both or neither of the white dwarfs are magnetic, than no current is set up. As the fraction of white dwarfs with magnetic fields is not known, it is difficult to speculate on the likelihood of this occurring. Finally, the radio emission may be beamed. We observe the orbit edge-on from Earth, so if the emission cone is too

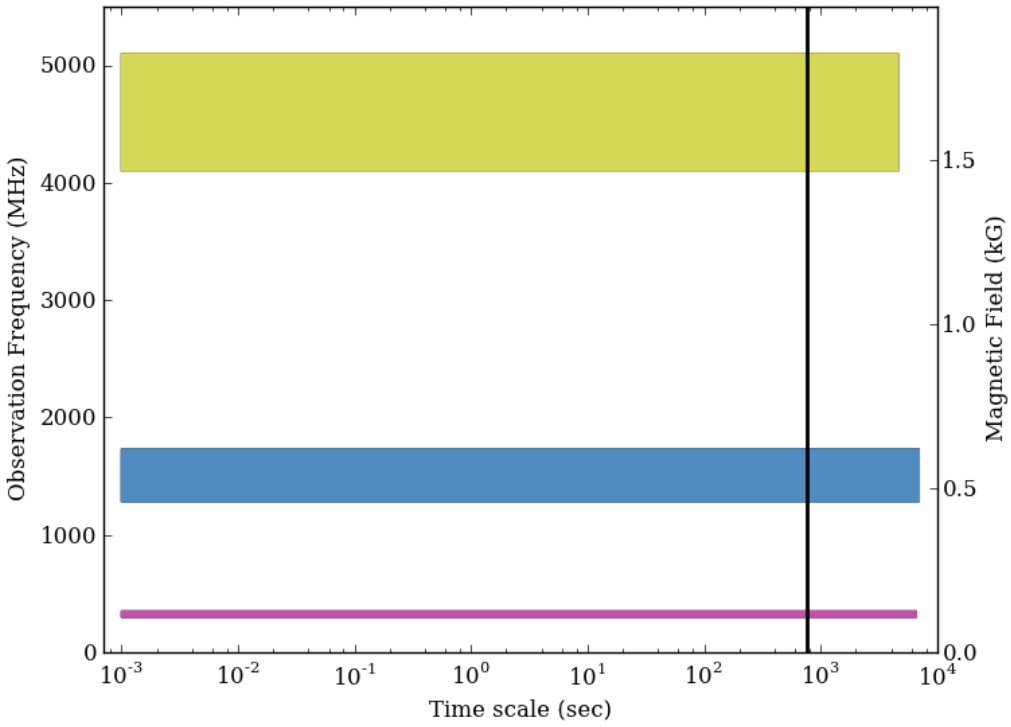


Figure 4.6: Binary white dwarf observational phase space plot. The observational frequency of our observations is shown versus the time scales of variability searched in our analysis. The magenta, cyan, and yellow boxes correspond to the 327 MHz, 1.5 GHz, and 4.6 GHz observations respectively. The solid vertical line denotes the binary's orbital period. Shown on the right is the magnetic field strength that would yield the emission frequency shown on the left if the emission is cyclotron ($\nu_0 = qB/2\pi m$).

narrow, it may not cross our line of sight.

## 4.5 Conclusion

We have observed the compact white dwarf binary J0651 with the Arecibo Observatory to search for variable radio emission. If one or both of the white dwarfs have a strong magnetic field, their close proximity opens up the possibility of magnetic interactions that may produce radio emission. Although we did not confirm the presence of short-duration, bursty emission or periodic emission modulated at the orbital period, these sources may still be variable on long time scales. This makes them possible targets for slow transient surveys for the next generation radio facilities, such as the VAST survey (Murphy et al., 2012) on ASKAP (Australian SKA Pathfinder), the ThunderKAT project on MeerKAT, or future surveys for slow transients with the Square Kilometer Array. A positive detection of radio emission from compact white dwarf binaries will likely require large bandwidths. Because of the large range of possible magnetic field strengths, a large bandwidth is needed to find the peak emission. Also, these sources are likely to be faint, so a wide bandwidth is also required to achieve the necessary SNR.

CHAPTER 5

# SMALL TELESCOPE SURVEY: A SURVEY FOR RARE, BRIGHT FAST RADIO TRANSIENTS

In this work we present results for a survey of bright, rare fast transients. The data were taken with the 3.8 m radio telescope on the roof of the Space Sciences Building with special-built fast dump digital FFT spectrometer. The survey collected data continuously in a drift scan mode over an elevation range of 20–90 deg. Three passes of this elevation range were made during a total of 105 days of observing. The data were searched for bright, single pulses using a pulsar search pipeline. We implement a new RFI identification algorithm based on the spectral modulation index (Spitler et al., 2012) and show that it is effective at reducing the amount of RFI in our data. No candidate dispersed pulse was seen. We set upper limits to the event rate of such pulses and compare this rate to other surveys.[1]

## 5.1 Introduction

The time variability of astronomical sources reveals much about the underlying astrophysics. At radio wavelengths transients generally arise from the interactions of charged particles in magnetic fields. Functionally transients are classified as either slow or fast depending on whether the effects of interstellar dispersion are important. For transient events shorter than about a second, the signal processing pipeline should include dedispersion. Pulsars, the most famous short radio transients, were discovered when the radio sky was first

[1] J. Cordes

observed with high time resolution (Hewish et al., 1968).

In the spectrum of pulsar variability there are two examples of short-duration transients: giant pulses Knight (2006) and rotating radio transients (RRATs, McLaughlin et al., 2006). We distinguish these from other forms of variability such as nulling (Gajjar et al., 2012) or intermittency (Kramer et al., 2006) by the optimal method of detection. Giant pulses and RRATs are best detected with single pulse techniques rather than periodicity searches (McLaughlin & Cordes, 2003). With the Arecibo Observatory a giant pulse can be detected once every $\sim$ 3 seconds at 1.4 GHz with observed flux densities as high as $10^3$ Jy (Cordes et al., 2004). Rotating radio transients are a subset of pulsars that emit pulses sporadically with rates[2] ranging from $\sim$ 1 - 100 pulses hr$^{-1}$.

Rare transients include a list of speculated sources, none of which have yet been conclusively proven. Hansen & Lyutikov (2001) suggest that a burst of radio emission may immediately precede the merger of two, in-spiraling neutron stars as their magnetospheres interact. They estimate a flux density of $\sim$ mJy for a system at a distance of $D_{\rm max}$ = 100 Mpc, which is in principle achievable with Arecibo. The estimated rate of binary neutron in-spiral is about 3 day$^{-1}$ Gpc$^{-3}$ (Kalogera et al., 2004; Lorimer et al., 2007). For a volume of $D_{\rm max}$ = 100 Mpc, this yields a rate of $\sim 3 \times 10^{-3}$ day$^{-1}$, well below the limits for any transient survey.

Rees (1977) suggested evaporating primordial black holes could produce a burst of radio emission generated by an expanding shell of electrons and positrons interacting with the interstellar magnetic field. Blandford (1977) then estimated the spectrum of such a burst.

Superconducting cosmic strings may be source of fast radio transients

[2] http://www.as.wvu.edu/~pulsar/rratalog/

(Vachaspati, 2008) across cosmological distances. Superconducting strings can oscillate and emit radio emission, and this emission can be burst-like if there are cusps or kinks in the string. Cai et al. (2012) derive estimates for the flux density and find the event rate of short-duration radio bursts with flux densities $\sim 10^3$ Jy from cusps and kinks to be 0.08 day$^{-1}$ sky$^{-1}$ and $3\times10^{-4}$ day$^{-1}$ sky$^{-1}$ respectively.

Several possible extragalactic dispersed pulses have been reported in the literature. The "Lorimer" burst was reported by Lorimer et al. (2007). It was found in pulsar survey data from the Parkes multibeam receiver at a position near the Large Magellanic Cloud. The dispersion measure of the pulse was too large to be accounted for by the electron column density in the galaxy and the LMC, and it was therefore projected to be extragalactic. Further analysis by Burke-Spolaor et al. (2011a) have since cast doubt on the original interpretation due to similar signals with characteristics more similar to RFI. Keane et al. (2012) report on a high-DM single pulse discovered in a reprocessing of the Parkes Multibeam Pulsar Survey data. While the galactic latitude of the pulse ($b = -4°$) could put it in our galaxy, the NE2001 model (Cordes & Lazio, 2002) can only about for $\sim 70\%$ of the dispersion measure, suggesting that this pulse may come from an extragalactic population.

In this paper we describe a survey for fast radio transient that uses a small, single-dish telescope. The advantage to this instrument is its large field-of-view. The survey was automated and provides a case study for an automated radio synoptic sky survey.

## 5.2 Transient Survey Parameters

In this section we discuss the efficacy of various radio astronomy facilities for transient detection and discuss the tradeoffs between observational and survey parameters. (See Cordes (1997) for a more detailed analysis.) We begin with the radiometer equation, which relates the observational quantity $S_{\min}$, the minimum detectable source flux density, to parameters describing the radio telescope and instrumentation

$$S_{\min} = \frac{mkT_{\text{sys}}}{A_e\sqrt{\Delta\nu\tau}} \tag{5.1}$$

where $m$ is the minimum SNR threshold, $k$ is the Boltzmann's constant, $T_{\text{sys}}$ is the system temperature, $A_e$ is the effective area of the telescope, $\Delta\nu$ is the data, and $\tau$ is the integration time. Note the most efficient way to lower $S_{\min}$ is to reduce $T_{\text{sys}}$ or increase the collecting area.

For simplicity, we will assume a survey is searching for a class of transient standard candles with bursts of luminosity $L_0$, width $W_{\text{t}}$, and event rate $\eta$. The maximum distance ($D_{\max}$) that such a source is detectable is given by

$$S_{\min} = \frac{L_0}{D_{\max}^2}. \tag{5.2}$$

Assuming an isotropic distribution of sources with a number density $n_{\text{s}}$, the total number of sources within the beam of a single pointing is proportional to the volume of space sampled, i.e.

$$N_{\text{s}} = \frac{1}{3}n_{\text{s}}\Omega_{\text{FOV}}D_{\max}^3 \tag{5.3}$$

where $\Omega_{\text{FOV}}$ is the instantaneous field of view. For a single dish telescope with a single pixel feed, this is simply given by $\Omega_{\text{FOV,i}} \sim \lambda^2/d^2$ where $\lambda$ and $d$ are observing wavelength and telescope diameter respectively. For a feed array or phased

array feed, the instantaneous field of view is $\Omega_{\rm FOV} = N_{\rm pixels}\Omega_{\rm FOV,i}$. Combining Equations 5.2 and 5.3, we get

$$N_{\rm s} = \frac{1}{3} n_{\rm s} \Omega_{\rm FOV} \left( \frac{L_0}{S_{\rm min}} \right)^{3/2}. \tag{5.4}$$

This equation shows that the most effective way to increase the volume is to decrease $S_{\rm min}$. If we also use the radiometer equation, we can expand Equation 5.4 to

$$N_{\rm s} = \frac{1}{3} \Omega_{\rm FOV} \left( \frac{L_0 A_e}{mkT_{\rm sys}} \right)^{3/2} (\Delta\nu\tau)^{3/4}. \tag{5.5}$$

Equation 5.5 tells the sampled volume increases most rapidly by increasing the collecting area or decreasing the system temperature. It is also clear that for a survey with a fixed total time, one samples more volume faster by observing over a larger field of view than with longer integration times.

Transient sources require additional factors to account for the event rate and finite duration. For this discussion we assume that $W_{\rm t} \ll T$ where $T$ is the duration of the pointing. For a drift scan survey the duration of a pointing can be taken to be the time a source needs to cross the FWHM beam width. The source volume density now also includes a rate factor: $n_{\rm t} = \eta n_{\rm s}$.

The number of transient sources observed in time $T$ is

$$N_{\rm t} = \frac{1}{3} n_{\rm t} \Omega_{\rm FOV} \left( \frac{L_0}{S_{\rm min}} \right)^{3/2} T. \tag{5.6}$$

It is useful to compare the relative numbers of sources detectable in two different surveys. If one survey detects a candidate fast transient, it is useful to understand what constraints on the event rate can be deduced from a different survey. Taking a ratio of Equation 5.6, we get

$$\frac{N_{\rm t,a}}{N_{\rm t,b}} = \frac{\Omega_{\rm FOV,a}}{\Omega_{\rm FOV,b}} \left( \frac{S_{\rm min,b}}{S_{\rm min,a}} \right)^{3/2} \frac{T_{\rm a}}{T_{\rm b}}. \tag{5.7}$$

A given survey places a constraint on the event rate ($R$) of sources with a flux density greater than $S_{\rm min}$ that occur once in the area sky observed during the entire survey.

$$R = \frac{N_{\rm t}}{\Omega_{\rm FOV,tot} T_{\rm pointing}} = \frac{N_{\rm t}}{\Omega_{\rm FOV} T_{\rm tot}} \tag{5.8}$$

In Section 5.4.1 we describe the event rate found with our survey and compare it to other fast transient surveys.

## 5.3 Survey Description

The Small Telescope Survey uses a 3.8 meter antenna located on the roof of the Space Sciences Building on the campus of Cornell University. The half power beam width (HPBW) of the antenna is about 4 deg. We observe in drift scan mode pointing south and step through an elevation range of ~20-90 degrees in increments of the half beam width (2 degrees). Each elevation position is observed for 24 hours, and then the telescope is automatically moved to the new elevation. A single pass of the sky takes 35 days and covers about 55% of the celestial sphere (~ 22,690 deg$^2$), and data are collected continuously. We have made three passes, which amounts to 105 days of observing. The exact dates of the three passes are given in Table 5.1.

| Start date | End date |
|---|---|
| 16 February 2011 | 24 March 2011 |
| 16 December 2011 | 22 January 2012 |
| 23 March 2012 | 5 May 2012 |

Table 5.1: Date ranges for the three passes of the sky.

The telescope is equipped with a 1.4 GHz feed. An analog filter selects a 57

MHz band centered at 1420 MHz. After mixing to an intermediate frequency of 260 MHz, the signal is sampled at 416 MHz by a CASPER[3] iADC+iBOB spectrometer (Roof Spectrometer). The digitized signal is mixed to baseband, and a 256-channel, 4-tap PFB is calculated. The complex FFT output is squared, and spectra are accumulated for 2 ms before being dumped over Ethernet. The observational parameters of the data are given in Table 5.2. Figure 5.3 is a diagram of the RF and IF bands, and the processing stages in the Roof Spectrometer.

Observations and data recording and analysis were controlled by a multi-threaded Python script. Newly acquired data were automatically moved to a Linux server for storage and sent to a local Linux computer cluster for processing. Upon the successful completion of the processing, diagnostic plots were created. The Python script also logged each raw data file and the processing status through a MySQL database.

| Property | Value |
|---|---|
| $T_{\mathrm{sys}}$ | 180 K |
| $\nu_{\mathrm{ctr}}$ | 1420 MHz |
| BW | 104 MHz |
| $\Delta\nu$ | 0.406 MHz |
| $N_\nu$ | 256 |
| $\Delta t_{\mathrm{s}}$ | 2 ms |

Table 5.2: System Parameters. Rows are as follows: $\nu_{\mathrm{ctr}}$ is center frequency of data, BW is spectrometer bandwidth, $\Delta\nu$ is frequency resolution, $N_\nu$ is the number of frequency channels, and $\Delta t_{\mathrm{s}}$ is the sample time of the spectra.

[3]https://casper.berkeley.edu/

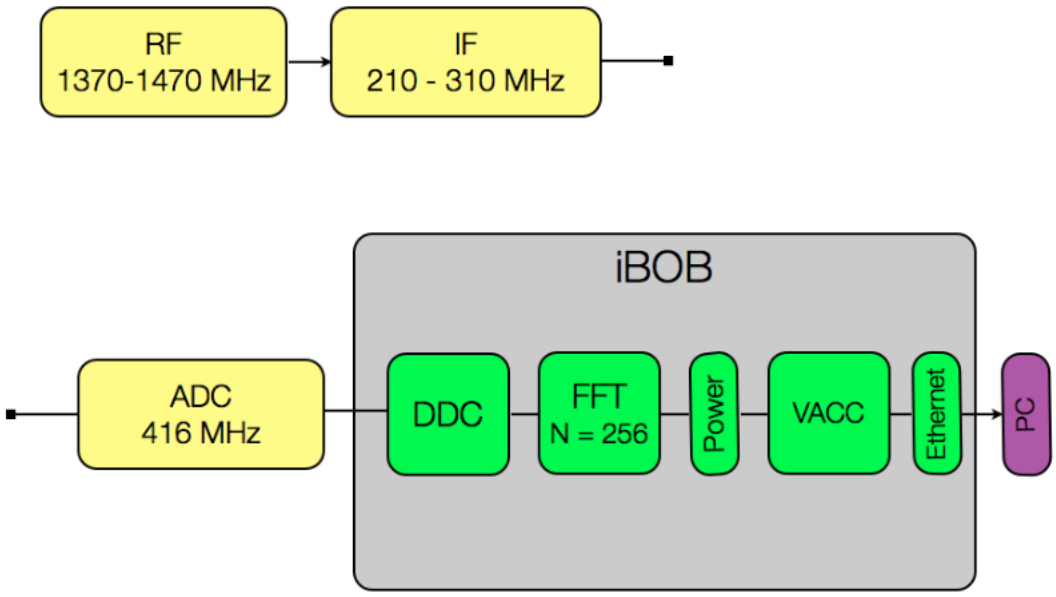


Figure 5.1: Diagram of the RF/IF bandwidths and Roof Spectrometer.

### 5.3.1 Data Processing

The data are processed using single pulse pulsar pipeline code developed at Cornell. The data are first dedispersed to generate dedispersed time series over a range of trial dispersion measures DM = 0 - 1000 pc cm$^{-3}$ in steps of 6 pc cm$^{-3}$. The code uses two algorithms for identifying individual pulses in the dedispersed time series. Both algorithms require candidate pulses to have a signal-to-noise ratio of $SNR_{min} > 5$ in the time series. The first algorithm applies a boxcar matched filter technique as described by Cordes & McLaughlin (2003). To increase the sensitivity to pulses that are broader than the native resolution of the data, multiple samples are averaged together. An efficient implementation averages adjacent samples in a dedispersed time series, applies the SNR threshold, decimates, and averages again. In our code this iterative smoothing

is performed seven times yielding boxcar widths of $2^n$ samples where $n = 0 - 7$ and effective time resolutions of $\Delta t_s = 2 - 256$ ms.

The second single pulse algorithm searches for temporal clusters in the candidate events defined as having signal-to-noise ratios above $\mathrm{SNR}_{\mathrm{min}}$. A set of events is consolidated into a cluster such that there is no gap in time greater than one sample. A cluster is characterized by statistics such as its average intensity, peak intensity, time of occurrence, and effective width. Unlike the boxcar smoothing algorithm, this technique is agnostic to pulse symmetry and can identify highly scattered pulses as easily as unresolved pulses.

For each trial dispersion measure, these two algorithms are applied to the dedispered time series. They generate a list of pulse candidates, which are inspected for characteristic signatures of a dispersed pulsar.

| Property | Value |
|---|---|
| DM range | 0 - 1000 pc cm$^{-3}$ |
| $\Delta$DM | 6 pc cm$^{-3}$ |
| $W_t$ | $2^n$ where $n = 0 \ldots 7$ |

Table 5.3: Pulsar pipeline parameters. Properties are as follows: DM range is the minimum and maximum trial dispersion measures, $\Delta$DM is the trial DM interval, $W_t$ is the width of the boxcar filters in sample units.

### 5.3.2 RFI rejection

Being located on a college campus, RFI is a major concern. Our primary RFI rejection technique is based on a spectral modulation index described in detail by Spitler et al. (2012). In brief, the spectral modulation index measures whether

a spectrum contains a broadband or narrowband signal. Any given candidate signal in a time series could arise from either a bright, narrowband signal or a moderately bright broadband signal, but once we average over frequency, we loose the ability to distinguish these two cases (without bringing outside information such as how SNR varies with DM). The spectral modulation index uses both the first and second moments of a spectrum, thereby retaining information on the distribution of the signal within the spectrum. Narrowband signals have a high modulation index, while broadband signals have a low modulation index. For example, a broadband signal in a spectrum from the Roof Spectrometer with SNR = 10 in the time series, has a modulation index of 1.3, while narrowband signal with a similar SNR has modulation index of ~13.

The calculation of the spectral modulation index can easily be automated and implemented in a software pipeline. Because the calculation requires the full time-frequency data, we take a two-pass approach. Candidate signals are first identified in a dedispersed time series as described above, and then we return to the appropriate segments in the time-frequency data to calculate the second moment. Spitler et al. (2012) show that for most typical survey parameters, this is more computationally efficient than calculating the second moment at the same time as the original calculation of the time series.

The modulation index provides a second discriminator of signal integrity. The first is SNR, which was applied in the original determination of a candidate list. Similarly a threshold in spectral modulation index ($m_{\mathrm{I,thr}}$) can be chosen to automatically classify of signals as either narrowband or broadband. In this survey our modulation index threshold is $m_{\mathrm{I,thr}} = 2.6$. We use the information provided by the modulation index as a discriminator of RFI, but in general it is

a tool to classify signals based on their spectral properties.

RFI excision occurs at two stages in our pipeline. First interference is identified in the time frequency plane before dedispersion. Then the candidate pulse list is sifted for events which are thought to be due to RFI. Steady, narrowband RFI was removed by identifying frequency channels that are anomalously high compared to the average bandpass. The worst type of interference in our data was broadband, impulsive RFI. Data were read in in 100 s chunks, and a time series was calculated by averaging over the frequency channels. For samples with an SNR > 5, the spectral modulation index was also calculated. A sample was considered too broad if its modulation index was less than $m_{\mathrm{I,thr}}$.

In our first pass, our data contained a very strong broadband signal that lasted N seconds and occurred every M seconds. These signals were not always thoroughly cleaned with the modulation index, so we also applied a cluster-based approach to identify regions that were particularly dense in strong events. When such a region was identified, it was removed entirely. Samples identified as bad either due to a low modulation index or as being part of a blast were replaced with noisy spectra. These noisy spectra contained Gaussian noise offset by the averaged bandpass and scaled by channel-by-channel standard deviation calculated using the clean spectra in the 100 second block. Note that we did not use the modulation index to identify narrowband RFI in the time-frequency data because a bright, broadband, dispersed pulse spanning several spectra will be narrowband in each of the spectra.

The spectral modulation index was used to identify narrowband RFI in the list of pulse candidates generated from our single pulse detection algorithms. Any candidate pulse with a modulation index greater than $m_{\mathrm{I,thr}}$ is removed from

the list. Figure 5.3.2 illustrates the effectiveness of the RFI rejection techniques. Each panel plots the dispersion measure of an event versus its time from the start of the file using an open circle with a size proportional to the SNR of the event.

The top panel shows the candidate pulses when no RFI rejection is done. The middle panel plots the resulting events when only RFI rejection is applied to the time-frequency data, and the bottom panel shows the events when RFI rejection is applied to both the time-frequency data and the list of candidates. The top panel is dominated by vertical stripes of events localized in time but spanning the entire DM range with SNR peaking near DM=0. This signature is typical of broadband, impulsive RFI. The middle panel illustrates that our modulation index-based broadband excision technique successfully removes these events. For this particular case only 14% of the original events are generated by the single-pulse algorithms. The remaining ~4 vertical stripes in the middle panel are characteristic of impulsive narrowband RFI (i.e. events spanning many trial DMs but with similar SNR). Applying our narrowband RFI rejection algorithm to the list of events further reduces the number of events by roughly a factor of six.

### 5.3.3 Analysis

The data were searched for astrophysical pulses by inspecting plots similar to Figure 5.3.2 for results from the matched filter and cluster algorithms seperately. A true, dispersed pulse has a well-defined signature in DM vs. time vs. SNR. The SNR peaks near the pulses's true DM. As the trial DM moves away from

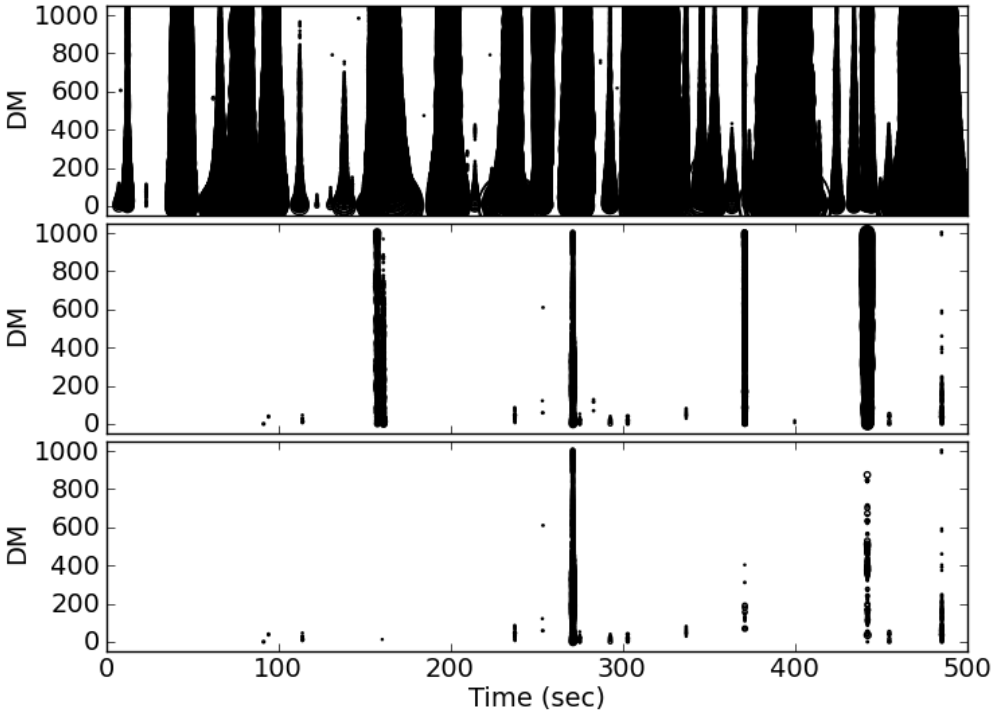


Figure 5.2: DM vs. time vs. SNR plot for 500 s of data from 17 February 2011 from the Small Telescope Survey. Each event in the candidate pulse list after various stages of RFI rejection are shown as open circles where the size of the circle is proportional to the event's SNR. The top panel illustrates the resulting events when there is no RFI rejection. The middle panel plots the events when RFI rejection is applied to the time-frequency data as described in Section 5.3.2, and the bottom panel shows the events when RFI rejection is applied to both the time-frequency data and the list of pulse candidates. The number of events in the three panels is 63277, 9138, and 1463.

the true DM, the SNR falls off and the location of the center of the pulse in time shifts to earlier and later times. Additionally, diagnostic plots such as SNR vs. DM, SNR vs. $m_{\rm I}$ and dynamic spectra were also generated. In the three passes no astrophysical pulses were discovered.

## 5.4 Discussion

### 5.4.1 Comparison with other Fast Transient Surveys

The minimum detectable flux density for the roof survey is $S_{\rm min}$ = 1.1 kJy assuming $W_{\rm t}$ = 1 ms and $m$ = 5. The upper limit to the rate of bursts with $S_{\rm min} > 1$ kJy is 30 day$^{-1}$ sky$^{-1}$. This is the most stringent limit for bright, rare transients. For context we compare these values to other surveys for fast transients in Figure 5.3 and Table 5.4. In each case we have assumed $W_{\rm t}$ = 1 ms and $m$ = 5.

The minimum detectable flux density and event rate limit for the Pulsar ALFA (PALFA, Cordes et al., 2006) survey operating at the Arecibo Observatory are taken from Deneva et al. (2009). This event rate represents about two-thirds of the total data collected to date, and the $S_{\rm min}$ quoted is for data collected with the Wideband Arecibo Pulsar Processor (WAPPs). The entire PALFA survey will cover low galactic latitudes $|b| \lesssim 5^\circ$ and galactic longitudes $30^\circ \lesssim l \lesssim 78^\circ$ and $162^\circ \lesssim l \lesssim 214^\circ$. PALFA-total represents the projected values for the completed survey assuming the total galactic longitude and latitude ranges have been observed once. We also used $S_{\rm min}$ for the Mock spectrometers, which have been in use since 2009 and have a three-fold increase in bandwidth over the WAPPs. In Table 5.4 we also give values for three smaller fast transient surveys

using Arecibo. Nice (1999) reports on a single pulse analysis of a narrow range of galactic longitude using an observing frequency of 430 MHz. Two surveys of the galaxy M33 have also been undertaken at 430 MHz (McLaughlin & Cordes, 2003) and 1.4 GHz with ALFA (Bhat et al., 2011).

Several single-pulse analyses have been reported from surveys conducted on the Parkes radio telescope. Data from the original Parkes Multibeam Pulsar Survey (PMPS, Manchester et al., 2001), which surveyed $50^\circ < l < 260^\circ$ and $|b| < 5^\circ$, have been searched by several authors for fast transient sources (McLaughlin et al., 2006; Keane et al., 2010). Two follow-up surveys covering the same galactic longitude as the PMPS but at higher galactic latitudes have been carried out. The Swinburne intermediate-latitude pulsar survey (Edwards et al., 2001) covered galactic latitudes $5^\circ \leq |b| \leq 15^\circ$, and a higher latitude extension (Jacoby et al., 2009) surveyed $15^\circ \leq |b| \leq 30^\circ$. Burke-Spolaor & Bailes (2010) reanalyzed these data for single pulses. Manchester et al. (2006) conducted a survey for pulsars in the Magellanic Clouds, and it was a reprocessing of the data that lead to the "Lorimer Burst" (Lorimer et al., 2007). The event rates for each survey individually is given in Table 5.4. To reduce clutter in Figure 5.3 , we only plot values for the original PMPS, which has the best constraints on the event rate and can be taken to be representative of the other Parkes surveys. The current pulsar and transient survey underway at Parkes is the High Time Resolution Universe (HTRU) pulsar survey (see Keane et al. (2010) for details). It will cover the entire sky visible at Parkes and is divided into three latitude ranges. Table 5.4 shows the event rates for each of the latitude ranges separately, as well as the combined event rate for the entire survey, which is also shown in Figure 5.3.

Wayth et al. (2012) report event rate limits from V-FASTR, a search for fast transients using the Very Long Baseline Array (VLBA) where each of the 25-m antennas is used as an independent station. This survey is different from the other efforts in that it is a commensal experiment and uses an array of dishes separated by large distances. The large physical separation between the antennas provides a powerful anti-coincident RFI rejection tool. One of the largest obstacles in identifying dispersed pulses is the large number of false positives due to RFI, and as Burke-Spolaor et al. (2011a) have pointed out, multi-dish facilities may be crucial to fully accepting individual single pulse candidates as real. As a commensal experiment, it has the potential to rapidly drive down the event rate limit through the availability of a large amount of observing time. Furthermore, V-FASTR searches for transients at all of the VLBA's frequency bands, where as most other surveys have been conducted near 1.5 GHz.

The Fly's Eye survey conducted with the Allen Telescope Array (ATA) (Siemion et al., 2012) achieved a low event rate by pointing the 42 ATA dishes at different positions in the sky. This yielded a large instantaneous field of view ($\sim 147$ deg$^2$), but the six meter ATA dishes give a modest $S_{\rm min}$.

### 5.4.2 FOV vs. Sensitivity

In Section 5.3 we outlined the relative importance of the sensitivity of a survey and the field of view by comparing the volume of space sampled. Equation 5.3 reveals that the volume is most sensitive to $S_{\rm min}$. Using Equation 5.7 we estimate that if the Small Telescope Survey had detected one, bright transient, we might expect the PALFA survey to have detected $\approx$ 86,000 fainter sources. This

| Survey | $S_{\min}$ (Jy) | Event rate sky$^{-1}$ day$^{-1}$ | Reference |
|---|---|---|---|
| Pulsar ALFA | $3.3 \times 10^{-3}$ | 93500 | Deneva et al. (2009) |
| Pulsar ALFA - completed | $1.9 \times 10^{-3}$ | 20812 | Deneva et al. (2009) |
| Arecibo 430 MHz survey | $70 \times 10^{-3}$ | 791139 | Nice (1999) |
| M33 survey - 430 MHz | $16 \times 10^{-3}$ | 952744 | McLaughlin & Cordes (2003) |
| M33 survey - 1.4 GHz | $3.3 \times 10^{-3}$ | 148148 | Bhat et al. (2011) |
| Swinburne inter. lat. | $17 \times 10^{-3}$ | 4790 | Edwards et al. (2001) |
| Swinburne high lat. | $17 \times 10^{-3}$ | 3333 | Jacoby et al. (2009) |
| PMPS | $17 \times 10^{-3}$ | 1029 | Manchester et al. (2001) |
| HTRU - high | $16 \times 10^{-3}$ | 661 | Keane et al. (2010) |
| HTRU - mid | $16 \times 10^{-3}$ | 1600 | Keane et al. (2010) |
| HTRU - low | $16 \times 10^{-3}$ | 1143 | Keane et al. (2010) |
| HTRU - total | $16 \times 10^{-3}$ | 332 | Keane et al. (2010) |
| V-FASTR | 0.5 | 9900 | Wayth et al. (2012) |
| Fly's Eye | 87 | 50 | Siemion et al. (2012) |
| Small Telescope Survey | 1100 | 30 | This work |

Table 5.4: Minimum detectable flux density and minimum event rate for fast transient surveys. The first column gives the name of the survey. The second column specifies the minimum detectable flux density (Jy) and minimum event rate (sky$^{-1}$ day$^{-1}$) for each of the surveys. The last column gives the reference for the values listed.

assumes an isotropic distribution of sources that extends to a distance greater than $D_{\max}$ for the more sensitive survey (i.e. uniform over cosmological scales). Instead if the sources are localized (i.e. Galactic) or not uniformly distributed in the sky (i.e. the Local Group), then the assumption of an isotropic distribution breaks down. In this case a system with a larger $S_{\min}$ and larger $\Omega_{\rm FOV}$ would be better-suited to detect such sources.

Galactic sources have a maximum distance that is truncated at the edge of the Galaxy. McLaughlin & Cordes (2003) show that Crab giant pulses could be detected to a distance of $\sim$ 1 Mpc in one hour of observing with Arecibo. Therefore Arecibo has the sensitivity to detect a Crab-like pulsar anywhere in the Galaxy if the only consideration was the reduction flux with increasing distance. Using a more sensitive telescope to probe to larger $D_{\max}$ will not help,

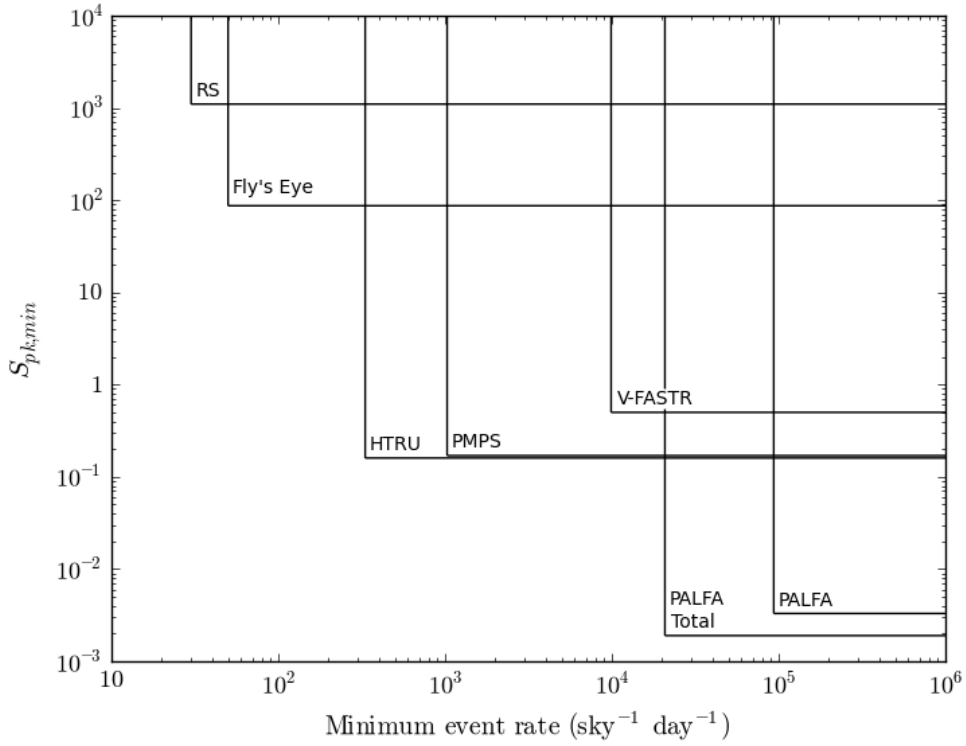


Figure 5.3: Minimum detectable flux density (Jy) vs. event rate (sky$^{-1}$ day$^{-1}$) for six fast transient surveys. A transient could be detected by a given survey if it lies inside the box for that survey. The survey labels and the parent telescope are as follows: Pulsar ALFA (PALFA), Arecibo; V-FASTR, VLBA; Parkes Multibeam Pulsar Survey (PMPS), Parkes; High Time Resolution Universe (HTRU), Parkes; Flye's Eye, ATA; Small Telescope Survey (RS). See Section 5.4 for references.

because one is probing beyond distribution of the population. In practice scattering limits our ability to detect a pulsar at large distances and at low Galactic longitudes.

The assumption of an isotropic distribution of sources also breaks down if the angular distribution of sources is non-uniform, for example in the Local Group. Stars, and presumably transient sources, are localized to galaxies. A

large FOV, and in particular a large survey speed, is needed to assure that the survey observes areas of the sky that contain sources.

From Equation 5.6 it is seen that the next best approach to increase the sampled volume is either observe longer or increase the instantaneous field of view. Increasing the observing span may not be feasible given proposal pressure and realistic time-scales for projects. The greatest scientific benefit then comes from increasing the field of view.

### 5.4.3 Transient Monitor

The Small Telescope Survey is a test-case for a radio transient monitor that continuously scans the sky for transient sources. Like the Small Telescope Survey, such a system would need to run autonomously for long periods of time.

Our experience also suggests that for a radio transient monitor to be successful, it needs a signal analysis pipeline that can identify real, transient signals and reject interference with high fidelity. In our case, the limiting factor for through-put was the manual inspection of the diagnostic plots. It would be entirely impractical for a large-scale monitoring effort to rely on humans to perform the initial pattern recognition and manually conduct follow-up. Great improvements need to be made in algorithms that classify RFI or astrophysical signals. The spectral modulation index as described in Section 5.3.2 is one example of such a technique. Spectral kurtosis is another technique in use by some other groups (Deller, 2010).

## 5.5 Conclusions

In this work we discuss the trade off between the minimum detectable flux density ($S_{\min}$) and the instantaneous field of view (FOV). The volume of space sampled by a telescope beam increases fastest with $S_{\min}$, but we also discuss situations where trading a higher $S_{\min}$ for a larger FOV is appropriate. The Small Telescope Survey is an extreme example of having a high $S_{\min}$ and large FOV. We conclude that such a system is too extreme, but that it is a useful test-bed for a more sensitive automatic transient detection system.

CHAPTER 6

# SEARCH FOR TRANSIENT SOURCES IN THE PULSAR ALFA SURVEY

We present first results from a single pulse analysis of data from the Pulsar ALFA survey (Cordes et al., 2006) operating at the Arecibo Observatory. Included in this paper are the first 23 months of data collected with the Mock spectrometers. A total of five new pulsars have been discovered and 47 previously known pulsars rediscovered. Three of the new discoveries are classified as Rotating Radio Transients (RRATs) with only a few individual pulses detected. The other two new pulsars were also found as periodicity candidates. The analysis includes several new algorithmic developments. First is the application of the spectral modulation index (Spitler et al., 2012) to automatically identify and reject events due to narrowband RFI. The second is a single pulse rating that identifies clusters in the dispersion measure - time plane that are consistent with the distribution of events for a real pulse. This rating is used to identify beams within pointings that are likely to contain pulsars. An analysis of the energetics of single pulse redetections of known pulsars reveals that approximately 28% of our redetections are consistent with a lognormal pulse amplitude distribution, which is consistent with the results from Burke-Spolaor et al. (2012).[1]

## 6.1 Introduction

The traditional "lighthouse" paradigm of pulsar emission evokes the picture of pulsars as steady radio emitters. In reality pulsar emission exhibits variations over a wide range of time scales. Roughly 100 pulsars exhibit nulling behavior where pulsation ceases abruptly for some number of periods. Each nulling pul-

[1]This work was conducted as a member of the Pulsar ALFA collaboration.

sar has its own characteristic nulling fraction and time scale that can be as short as several periods to as long as months (Backer, 1970; Wang et al., 2007; Gajjar et al., 2012; Kramer et al., 2006). The pulsars which exhibit nulls over the long times scales (10s - 100s of days) are often called intermittent pulsars. Changes in the emission properties are linked to changes in the pulsar magnetosphere that have been found to correlate with the pulsar spin down rate. Kramer et al. (2006) found that the intermittent pulsar J1933+2421, is "on" for five to 10 days and "off" for 25 to 35 days, and has a spin down rate 50% higher during the "on" state than the "off" state. J1841-0500 is another highly intermittent pulsar for which Camilo et al. (2012) compared the "on" and "off" spin down rates after a 580-day null and found that the spin down rate was 250% higher in the "on" state.

The subclass of pulsars dubbed RRATs (Rotating Radio Transients; McLaughlin et al., 2006) exhibit a complementary type of variability in which only a few, often isolated pulses, are emitted. RRATs are best found in surveys through single pulse techniques which have become ubiquitous in pulsar survey pipelines (Cordes et al., 2006; Deneva et al., 2009; Burke-Spolaor et al., 2011b). The periods of the known RRATs are consistent with the periods of canonical pulsars but tend to be slow rotators with periods of $\sim$ 0.5 - 8 s (Keane & McLaughlin, 2011).

Single pulse search techniques applied to high time resolution pulsar survey data may also detect radio emission from new populations of rare radio transients that to date have been mere speculation. A burst of short duration radio emission may immediately precede the merger of two, in-spiraling neutron stars (Hansen & Lyutikov, 2001). Rees (1977) suggested that one could probe the

population of primordial black holes using bursts of radio emission produced when they evaporate. Kinks or cusps in superconducting cosmic strings may generate radio bursts (Cai et al., 2012). The primary limiting factor in detecting these rare sources is not only sensitivity, but also the rate at which the sky is observed. New technologies such as phased array feeds or aperture arrays will be crucial to detecting these rarer transient sources.

## 6.2 Survey Description

The Pulsar ALFA (PALFA) survey is searching for pulsars in the galactic plane and has been running at the Arecibo Observatory since 2004. The survey uses the Arecibo L-band Feed Array (ALFA[2]), a seven-pixel receiver housed in the Gregorian dome. The pixels are arranged with a single center pixel (Beam 0) surrounded by a hexagonal ring of six pixels (Beams 1 – 6). The FWHM beam size of each ALFA pixel is 3.5 arcmin. The RF frequency range of ALFA is 1225 - 1525 MHz and has a typical system temperature of 30 K. The on-axis gain Beam 0 is 10.4 K $Jy^{-1}$, and the average on-axis gain of the other six beams is 8.2 K $Jy^{-1}$. The peak gain of the sidelobes (about $\sim 5$ arcmin from beam center) is 1.7 K $Jy^{-1}$.

For a full description of the PALFA survey parameters, see Cordes et al. (2006); we summarize the relevant properties here. The survey observes at low galactic latitudes ($|b| \leq 5°$) and is broken into two ranges of galactic longitude: inner galaxy ($30° < l < 78°$) and outer galaxy ($162° < l < 214°$). The duration of a single pointing typically is 268 s. Some outer galaxy pointings have durations of 134 s, and the data collected in January through May 2010 have durations of

[2] `http://www.naic.edu/alfa/`

512 s.

From the onset of observing until Spring of 2009, the survey used the Wideband Arecibo Pulsar Processor (WAPPs) for data collection. The WAPPs were limited to a bandwidth of 100 MHz, less than what is available with ALFA. In March of 2009, PALFA began observing with the Mock/PDEV[3] spectrometers (hereafter, Mock spectrometers). The Mock spectrometers cover the entire ALFA frequency range in two subbands of 172 MHz each. These two subband are merged by the software pipeline, resulting in 322.6 MHz of total bandwidth centered on 1375.4 MHz with a frequency resolution of 0.34 MHz. The time resolution is 65.5 $\mu$s. The additional bandwidth provides an increase in sensitivity of up to 70%, but radio frequency interference (RFI) can reduce the amount of usable bandwidth.

Deneva et al. (2009) describe the single pulse analysis and results from the WAPP data. In this analysis we focus on data from the Mock spectrometers and present results of the observations from March 2009 through January 2011. These data comprise a total of 2259 pointings, of which 1778 are inner galaxy pointings and 481 are outer galaxy pointings. Figure 6.1 illustrates the locations of these pointings in Galactic coordinates split into the inner Galaxy (top) and outer Galaxy (bottom). The total area of sky covered is 42 deg$^2$ (inner galaxy: 33 deg$^2$, outer galaxy: 9 deg$^2$).

[3] http://www.naic.edu/~phil/hardware/pdev/usersGuide.pdf

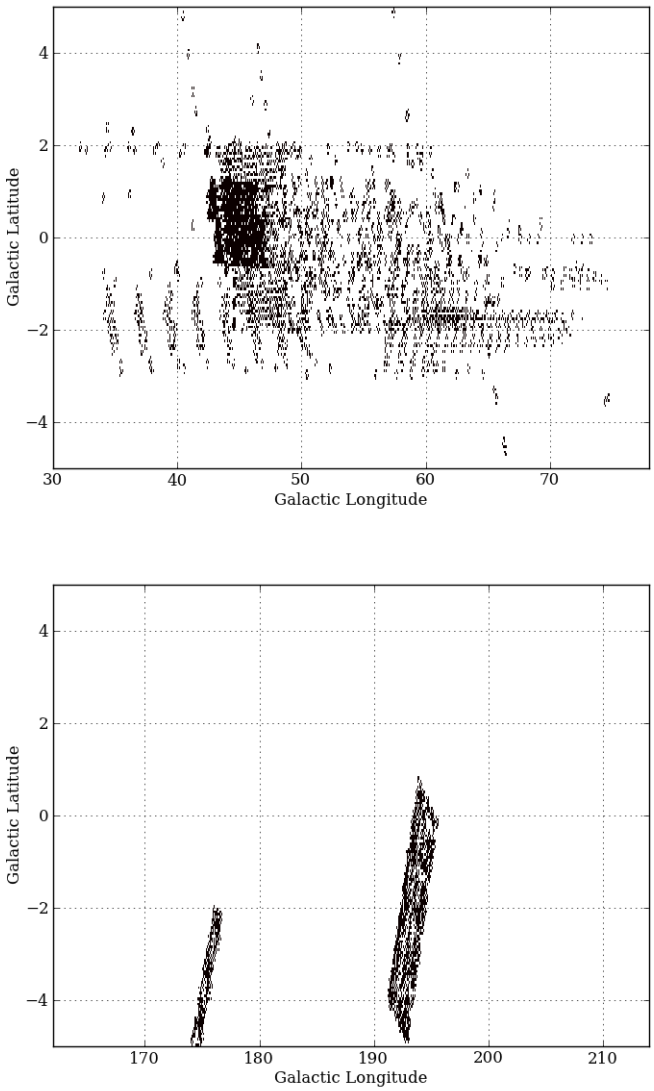


Figure 6.1: Galactic coordinates for the PALFA pointings presented in this work. The top and bottom panels show the inner and outer Galaxy pointings respectively.

## 6.3 Single Pulse Analysis

The time-frequency data are processed with a PRESTO[4]-based pipeline to search for single dispersed pulses. The raw data are cleaned of RFI using the standard PRESTO RFI excision code (rfifind). Periodic, narrowband RFI is identified by calculating a Fourier transform along the frequency channels and reporting channels with power above a threshold. Impulsive RFI is removed by comparing the statistics of each spectrum to the statistics of a few second-long block of data. Time-frequency samples identified as RFI are replaced values representative of the statistics of the rest of the data.

The raw time-frequency data are dedispersed with 5016 trial dispersion measures ranging from DM = 0 – 2038 pc cm$^{-3}$. As the trial DM increases, the step size between trial DMs and the effective time resolution increase. The effective time resolution is the quadratic sum of several factors including the native time resolution of the data and residual dispersion smearing. The dominate effect is the residual DM smearing within a single frequency channel. Eventually the residual DM channel smearing is larger than the time resolution of the data, and the minimum pulse width is more than one sample wide. The native resolution of the data is now too large, and one can reduce the computational burden of dedispersion by averaging samples to match the DM channel smearing. At our highest DMs, the time resolution is reduced by a factor of 15. Since the processing requirements of dedispersion are proportional to the number of samples, this brings significant computational savings.

Pulse candidates are identified with two algorithms applied to the dedispersed time series: smoothing and time domain clustering. The smoothing al-

[4] http://www.cv.nrao.edu/~sransom/presto/

gorithm increases the sensitivity to pulses wider than the time resolution of the data by convolving the dedispersed time series with a series of boxcar matched filters. (See Cordes & McLaughlin (2003) for a general discussion of this technique.) The widths of the filters range from 2 - 1500 time samples making the effective time resolution for the widest filter $\sim$ 1 s at low DMs and $\sim$ 15 s at high DMs. After convolution a threshold signal-to-noise ratio (SNR) > 5 is applied. A bright signal in a time series may be detected at multiple filter widths, so the results are sifted to return only the candidate event associated with the width that yields the largest SNR. A boxcar filter is the obvious choice of filter shape because of its simplicity, but is not necessarily the optimal choice for real pulse shapes, particularly for a highly scattered pulse that is asymmetric with a roughy exponentially decreasing tail.

Our second algorithm is a time-domain clustering, or "friends-of-friends", technique and is agnostic to the pulse shape. This algorithm is applied to the list of candidate events generated by applying the SNR threshold to the full-resolution time series. Multiple events are consolidated into clusters under the criterion that the gap between any two consecutive events no more $n_{\rm gap} = 1$ sample in our analysis. The disadvantage of technique is a pulse must be bright enough to be detected in the full resolution time series without the sensitivity benefits of smoothing.

Dedispersion combines frequencies according to the cold plasma dispersion law and increases the SNR in the time series by up to $\sim \sqrt{N_\nu}$. A bright signal in a time series may or may not be broadband and by averaging information about the distribution of power across frequency is lost. But by also calculating the second moment, we can regain some of this information. Specifically we use the

spectral modulation index, described in detail in Spitler et al. (2012), to estimate how broadband or narrowband a signal is. Spectra that are broadband have low modulation indices, and spectra that are narrowband have high modulation indices. For each candidate event produced by our single pulse algorithms, we test its modulation index against a modulation index cutoff to automatically classify as broadband or narrowband; those deemed too narrowband removed from further analysis.

We look for dispersed pulses in the list of candidate pulses on a pointing-by-pointing basis with seven-panel diagnostic plots similar to Figure 6.2. Each horizontal frame displays the candidate events for each of the seven ALFA beams with Beam 0 on the bottom and Beam 6 on the top. The main panel plots each candidate vs. DM and time with an open circle whose size is proportional to SNR. The middle subplot shows SNR vs. DM and the right-most subplot is a histogram of SNR. To avoid clutter, we generate plots over three DM ranges. Figure 6.2 is from a data set that contains a known pulsar: B1924+16. There is a train of single pulses at DM = 177 pc cm$^{-3}$, and a peak in the SNR vs. DM panels. The isolated events at non-zero DMs are caused by weak RFI or are false positives from noise.

### 6.3.1 Single Pulse Rating

Traditionally single dispersed pulses are identified in a list of candidate pulses by manually inspecting plots like Figure 6.2. This process is time consuming and subject to human error. If a large scale effort to search for dispersed pulses is going to succeed, algorithms that can automatically identify pulse candidates

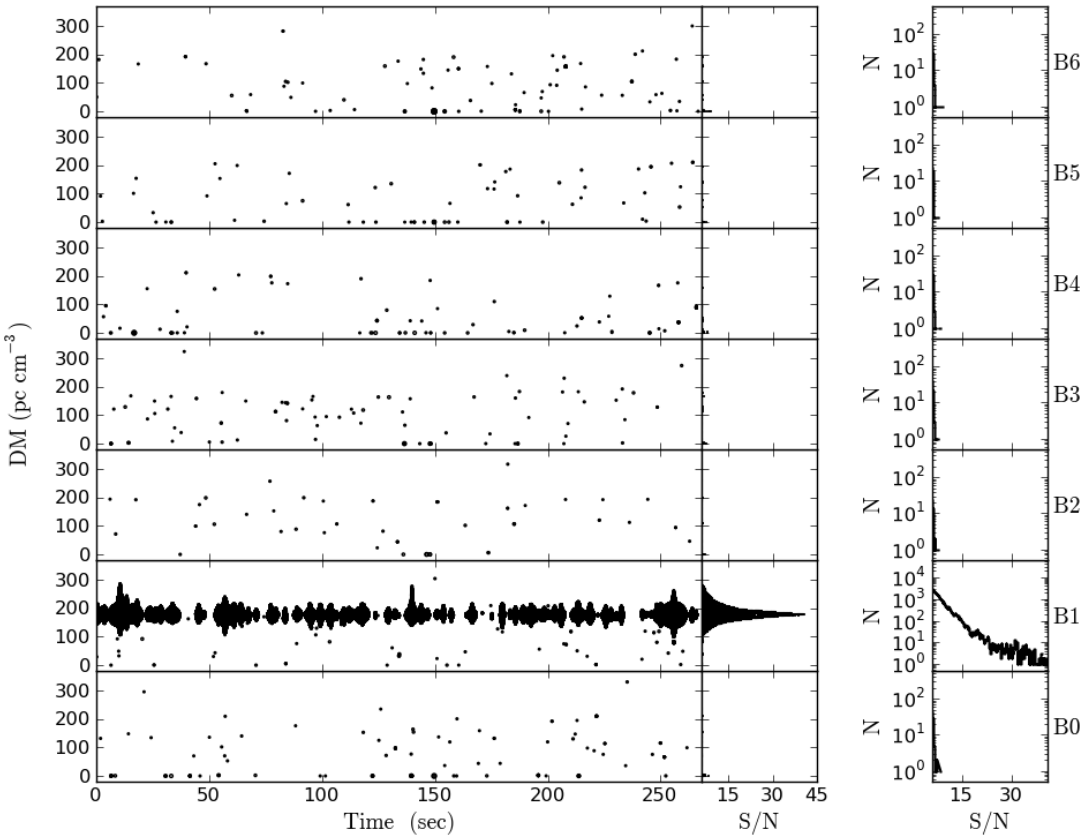


Figure 6.2: Seven panel single pulse diagnostic plot. Pulsar B1924+16 is visible in Beam 1 near DM = 177 pc cm$^{-3}$. The pulsar pulses produces a peak in the SNR vs. DM subplot and fills the SNR histogram to higher values.

must be developed.

It is common for pulsar periodicity searches to use rating schemes to select promising pulsar candidates. Several pipelines feed statistical properties into artificial neural networks to synthesize these properties into a rating (Eatough et al., 2010). We discuss the development of a single pulse rating scheme which can help identify possible dispersed pulses within an event list. In our analysis we applied the rating algorithm to each beam in a pointing and quantified the

number of pulse-like clusters in that beam.

The characteristic property of a dispersed pulse is that its SNR is a maximum at the correct dispersion measure, and as the trial DM moves away from the correct value, the SNR of the pulse drops until it falls below threshold. This occurs because a DM error leaves the pulse with some residual time sweep that has not been properly accounted for. Residual DM time smearing also has the effect of shifting the center of the pulse to earlier or later times. Let the arrival time of the peak of a pulse be $t_0$ and have a true dispersion measure of $\mathrm{DM}_0$. A dispersion error $\delta\mathrm{DM} = \mathrm{DM} - \mathrm{DM}_0$ will result in the pulse being smeared in time by

$$\delta t_{\mathrm{DM}} = 8.3\mu\mathrm{s}\ \delta\mathrm{DM}\frac{\Delta\nu_{\mathrm{MHz}}}{\nu^3_{\mathrm{GHz}}}. \tag{6.1}$$

where $\delta t_{\mathrm{DM}} = t - t_0$, $\Delta\nu_{\mathrm{MHz}}$ is the data bandwidth in MHz, and $\nu^3_{\mathrm{GHz}}$ is the observing frequency in GHz. For PALFA data taken with the Mock spectrometers, a DM error of $\delta\mathrm{DM} = 1$ pc cm$^{-3}$ results in a smearing of $\sim 1$ ms.

Figure 6.3 shows the shift in pulse arrival time with increasing DM error for a bright pulse from pulsar J1910+05. For clarity the true DM and arrival time of the pulse, defined to be the event with the maximum SNR, has been subtracted off to show the DM and arrival time offset. Events at DMs higher than the correct DM drift to earlier times while events at DMs lower than the correct DM drift to later times. This particular slope in DM–Time is valid for dedispersion code referencing the arrival time to the top of the frequency band. For dedispersion code that references the pulse arrival time to the lowest frequency in the band, the temporal drift is flipped, but the following discussion is still valid. Dedispersion code that references to the band center results in no drift, and the following rating would not work without modification.

This predictable drift in arrival time and trial DM provides a simple test for how pulse-like a cluster of events in DM and time is. Specifically, half the events in Figure 6.3 lie in the upper left quadrant (Q0), and half the events lie in the lower right quadrant (Q3). The fraction of the total number of events in QN is given by $f_{\rm QN} = N_{\rm QN}/N_{\rm tot}$. We then define a single pulse rating in terms of the fraction of events in each of the four quadrants:

$$R_{\rm sp} = 1 - |0.5 - f_{\rm Q0}| - |0.5 - f_{\rm Q3}| - f_{\rm Q1} - f_{\rm Q2} \tag{6.2}$$

An ideal pulse has Q0 = Q3 = 0.5, Q1 = Q2 = 0, and $R_{\rm sp} = 1$. The rating for the pulse shown in Figure 6.3 is $R_{\rm sp} = 0.94$. The deviation from 1.0 is due to a few events drifting into Q2 at small $\delta$DM. Empirically we found that real dispersed pulses have ratings $R_{\rm sp} > 0.75$ and used this value as a threshold to define clusters that are highly pulse-like.

The size of the box depends on the SNR of the central event, because a brighter pulse can be detected to larger DM errors than a weaker pulse. The drop off in SNR with $\delta$DM can be estimated using Equation 6.1. The exact relationship of SNR vs. DM depends on details such as the pulse shape, pulse width, and matched filter width, but we have found that the algorithm is not strongly dependent on the details of the box size.

This quadrant-based rating is well-suited to reject clusters at low DM caused by broadband RFI. A broadband RFI signal undergoes the same drift in DM and time as non-zero DMs impose a negative dispersion on the spectrum. The key difference is the peak SNR occurs at $\rm DM_0 = 0$ pc cm$^{-3}$ and trial DMs are chosen to be only positive, so the DM–arrival time drift is not symmetric. But if we draw a box around a bright event at DM = 0 pc cm$^{-3}$, events will only fall in Q0. For this maximally asymmetric case, $f_{\rm Q0} = 1$ and $f_{\rm Q1} = f_{\rm Q2} = f_{\rm Q3} = 0$, which yields

$R_{sp} = 0.0$. Also note that a uniform distribution of events in all four quadrants (i.e. $f_{Q0} = f_{Q1} = f_{Q2} = f_{Q3} = 0.25$) results in $R_{sp} = 0$.

While $R_{sp}$ reliably gives real, dispersed pulses a high rating, some RFI can also be rated with a high $R_{sp}$. In the PALFA data events associated with the airport radar often pass the pulse rating threshold but are clearly not dispersed pulses when viewed in a plot similar to Figure 6.3. The airport radars have a period of 12 s, and each pulse is comprised of shorter duration sub-pulses. For clusters which pass the first pulse rating test, i.e. $R_{sp} > 0.75$, we apply a second test.

The second test relies on the fact that a real dispersed pulse should be detected over a range of trial DMs from the true value to the DM where the residual time smearing reduces the SNR of the pulse below the detection threshold. For the pulse shown in Figure 6.3, the trial DM spacing is 0.3 pc cm$^{-3}$, and it is clear that there are events at each trial DM until the edges of the box. Knowing the size of the box, the number of expected events can be estimated and compared to the number of actual events. Specifically a DM filling factor is defined as $f_{DM} = N_{events}/N_{expected}$. The average DM filling factor for Q0 and Q3 in Figure 6.3 is $f_{DM} = 0.95$. Empirically we found that a DM filling factor $f_{DM} > 0.75$ reliably returned real pulses while rejecting most of the RFI.

We apply these pulse recognition algorithms to the list of single pulse candidates generated for each beam of a single pointing to estimate the number of pulse-like clusters. The algorithm is applied iteratively. In each iteration, the event with the maximum SNR is identified and a box like the one shown in Figure 6.3 is drawn around this event. The single pulse ratings $R_{sp}$ and $f_{DM}$ are calculated. If the cluster has $R_{sp} > 0.75$ and $f_{DM} > 0.75$, the value of $R_{sp}$ and the

DM of the cluster is recorded. is added to a cumulative rating as an estimate of the number of pulse-like clusters . All the events inside the box are removed from the list of candidate events, and a new iteration begins by finding the event with the highest SNR among those that remain. This process is repeated until all of the events with SNR > 7 have been removed from the list.

Once the loop finishes, a cumulative pulse rating ($N_{\rm sp}$) is calculated by summing the $R_{\rm sp}$ values for all of the pulse-like clusters. This gives an estimate of the number of pulses in an observation but weighted by the quality of the pulse. Additionally, the mean and standard deviation, $\rm DM_{mean}$ and $\rm DM_{std}$, of all of the clusters are also calculated to determine whether the pulses have a similar DM, as would be expected for a pulsar, or a wide range of DMs, as would be expected for RFI. The major drawback to this simplistic metric is if a pulsar and strong RFI exist in the same pointing then the mean may not properly reflect the DM of pulsar and standard deviation would be too large. A more sophisticated metric could histogram the DMs of the pulse-like clusters and report the bin with the largest number of pulses.

The pulse metrics for the Beam 1 of Figure 6.2 are $N_{\rm sp} = 110.4$, $\rm DM_{mean} = 177.0\ pc\,cm^{-3}$, and $\rm DM_{std} = 2.8\ pc\,cm^{-3}$. These metrics then tell us that there are roughly 110 pulses in Beam 1 with a DM near 177 $\rm pc\,cm^{-3}$ and SNR > 7. The number of pulse-like clusters identified for the other six beams is 0. In Section 6.4 we give $N_{\rm sp}$, $\rm DM_{mean}$, and $\rm DM_{std}$ for our discoveries.

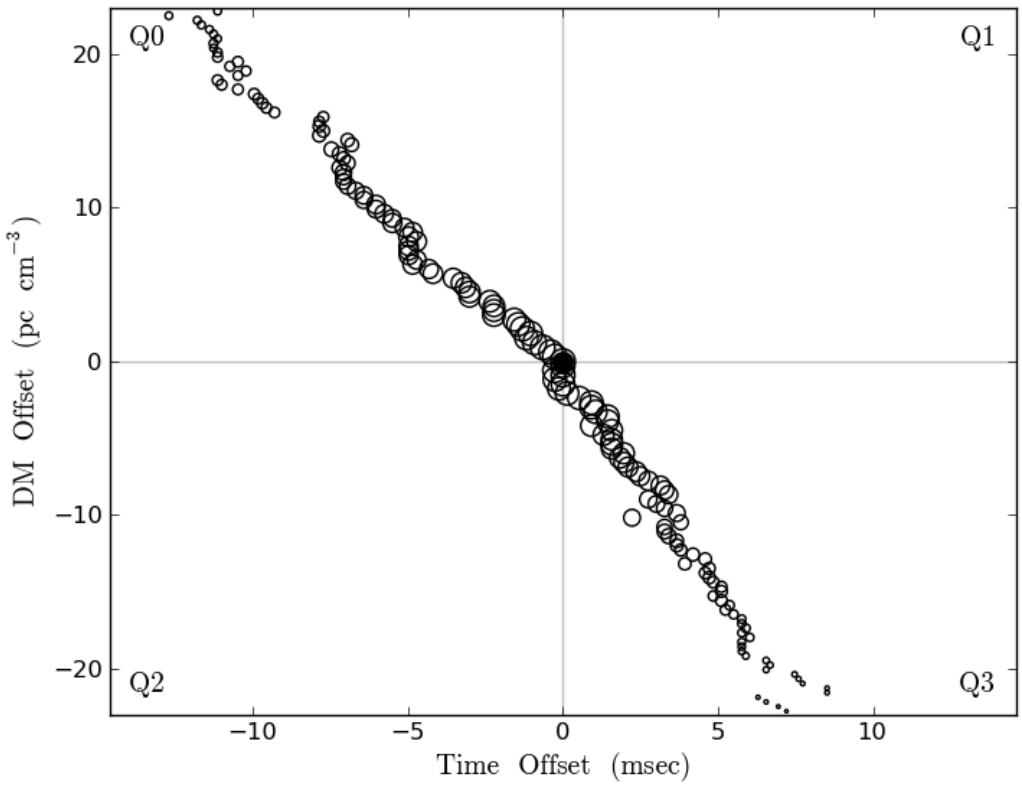


Figure 6.3: DM vs. Time vs. SNR plot illustrating the in time for a real dispersed pulse due to dedispersion with an incorrect DM. The size of the circles scale with SNR. The event with the highest SNR is shown with a filled circle. The effect of residual DM error causes a drift in pulse arrival times to earlier and later times.

## 6.4 Discoveries

In this section we briefly describe the five single pulse discoveries made in the reprocessing of the PALFA data. The properties of these pulsars are summarized in Tables 6.1 and 6.2. Two pulsars, J1928+14 and J1905+10, were also detected in a periodicity analysis results from the Einstein@home pipeline using the parameters of the single pulse discovery. The Einstein@Home distributed

volunteer computing project uses idle compute cycles on home and/or office PCs donated by members of the general public. The long-term and main goal of Einstein@Home is the detection of gravitational waves from rapidly rotation neutron stars in data from ground-based interferometric detectors. A fraction of the available computing power is used to search data from the PALFA pulsar survey with Arecibo for pulsar in very compact binaries, i.e. with orbital periods as short as 11 minutes.

The columns of Table 6.1 are as follows: pulsar name in J2000 coordinates, right ascension, declination, period, pulse width in ms, dispersion measure, and pulse flux density. The columns of Table 6.2 are as follows: pulsar name in J2000 coordinate, length of the observation in seconds, number of pulses observed, pulse occurrence rate, and section reference. For each pulsars we report the width of the brightest pulse ($W_{\rm t}$) corresponding to the boxcar filter width that yielded the highest SNR, and for the pulsars with periodicity detections, the full width half maximum of the average pulse profile ($W_{50}$) is reported in parenthesis. Similarly, the flux density of the brightest pulse is reported as the peak flux density ($S_{\rm p}$) for all objects, and the flux density ($\langle S \rangle$) of the pulse profile averaged over the full period is reported for the pulsars with periodicity detections.

### 6.4.1 J1929+11

J1929+11 was discovered in observations from 17 March 2009. A total of four pulses with SNR > 6 at a DM = 80.5 pc cm$^{-3}$ were detected; the brightest with a SNR $\approx$ 11. The top panel of Figure 6.4 shows the four pulses in a DM versus time plot from the discovery observation. The real pulses are denoted with ar-

rows. The other events at similar DMs were thoroughly checked and could not be confirmed to be pulses. We presume that they are weak RFI. The four pulses follow an underlying period of 3.218 s, but the true period may be shorter multiple of this value. Each pulse was verified to be broadband and dispersed after inspection of the time-frequency data. The bottom panel of Figure 6.4 shows the brightest pulse in the time-frequency data. The vertical yellow stripes are the regions where bright, broadband RFI has been removed. The brightest two pulses were discovered with the matched filter algorithm at the same pulse width, so we estimate the width of the pulses to be 1.3 ms.

The single pulse ratings for the discovery observation was $N_{sp}$ = 1.0, $DM_{mean}$ = 80.4 pc cm$^{-3}$, and $DM_{std}$ = 0.0 pc cm$^{-3}$. These statistics suggest the presence of a single pulse with SNR > 7 and a DM near 80 pc cm$^{-3}$. This matches our observation of one pulse at SNR $\approx$ 11 and three pulses with 6 < SNR < 7.

Follow-up observations of J1929+11 were conducted on 19 September 2012 with the PUPPI instrument at the Arecibo Observatory. The L-wide receiver was used with a total bandwidth of 800 MHz, 2048 frequency channels, and a time resolution of 40.96 $\mu$s. In the ten minute observation, no dispersed pulses were seen. If the RRAT were pulsing at the same rate as in the discovery observation, we would have expected to see 32 pulses with the sensitivity of the Mock spectrometers. The absence of pulses suggests J1929+11 may be intermittent on long time scales. Alternatively, the wider bandwidth of L-wide results in more RFI in the band, and we may be not be properly removing it.

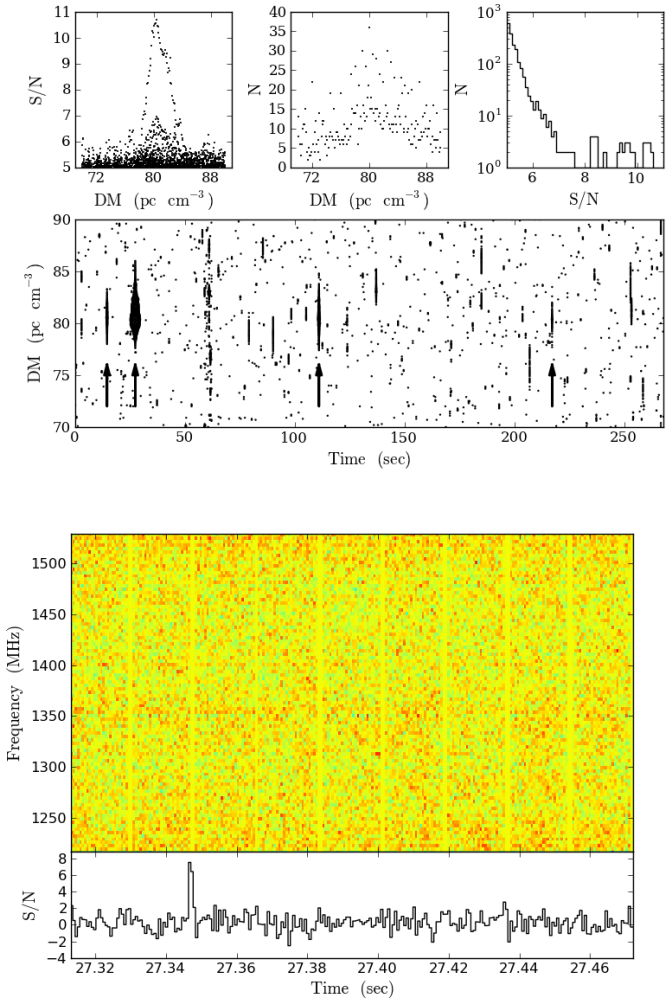


Figure 6.4: Discovery plots of J1929+11. The top four panels show the single pulse diagnostic plots for the discovery observations over a narrow range of DMs centered on the true DM of the pulsar. The four pulses are near $t = 14, 27, 110$ and 217 and denoted with arrows. The bottom panel shows the time frequency data for the brightest pulse; the dispersed pulse is clearly visible. The vertical stripes are regions where broadband RFI has been removed. The lower panel shows the dedispersed time series.

### 6.4.2 J1913+06

J1913+06 was discovered in observations from 15 April 2009 in which seven pulses with SNR > 5 were detected. The two brightest had SNR ≈ 9 and an average width of about 2 ms. The top panel of Figure 6.5 shows the DM vs. time plot from the discovery observations. The seven pulses have an underlying period of 1.257 s are located near t = 47, 62, 64, 67, 101, 145, and 234 s and are denoted with arrows. The other signals near the DM of the pulses are RFI. The bottom panel of Figure 6.5 shows the time-frequency data for the pulse near t = 67 s. The single pulse rating statistics are $N_{\mathrm{sp}} = 1.9\mathrm{pc\ cm^{-3}}$, $\mathrm{DM_{mean}} = 143.9\mathrm{pc\ cm^{-3}}$, and $\mathrm{DM_{std}} = 0.5\mathrm{pc\ cm^{-3}}$. The rating accurately predicted two pulses with SNR > 7 and the correct DM.

### 6.4.3 J1928+14

J1928+14 was discovered in observations from 21 June 2010 as a single pulse candidate and then also identified as a periodicity candidate in results from the Einstein@home pipeline. RFI in the data caused the folding step in the Einstein@home pipeline to optimize the pulsar's DM to 0 pc $\mathrm{cm^{-3}}$, and it was subsequently rejected as a candidate. When the data were folded at the DM and period determined from the single pulse discovery, it yielded a clear detection.

Twenty pulses were found with SNR > 5 and are shown in the top panel of Figure 6.6. The DM and period of the pulsar as determined by the periodicity analysis is 104.2 pc $\mathrm{cm^{-3}}$ and 1.0109 s. For this pulsar we report both single pulse properties and periodicity properties in Table 6.1. The lower panel of Figure 6.6 shows the time-frequency data for the brightest pulse. The single pulse

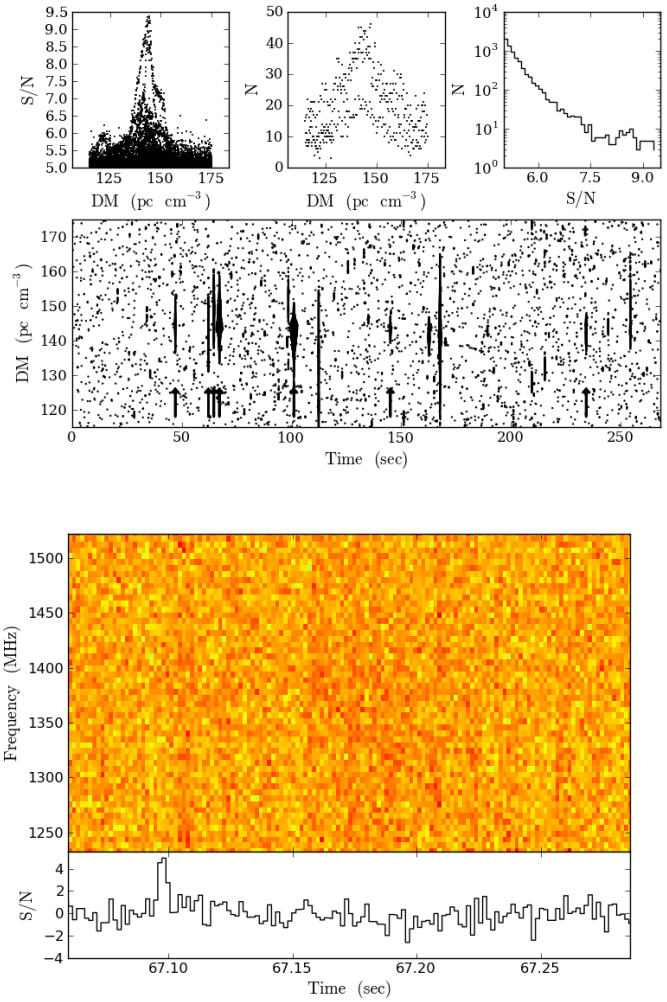


Figure 6.5: Discovery plots of J1913+06. The top four panels show the single pulse diagnostic plots for the discovery observations over a narrow range of DMs centered on the true DM of the pulsar. The seven pulses are near t=47,62,64,67,101,145 and 234 and denoted with arrows. The bottom panel shows the time frequency data for the brightest pulse located near t=67 s, as well as a dedispersed time series.

rating statistics for this discovery are $N_{sp}$ = 6.2, $DM_{mean}$ = 105.2pc cm$^{-3}$, and $DM_{std}$ = 0.6pc cm$^{-3}$. The number of pulses with SNR > 7 in this observation was twelve. The pulse rating under predicts the number of pulses in this observation. More than ~ 6 pulses were may have been identified by the algorithm but with moderately high pulse ratings, so the cumulative rating was lower. The rating also slightly over predicted the DM compared to the periodicity analysis.

### 6.4.4 J1905+10

J1905+10 was discovered in observations from 19 August 2010 as a single pulse candidate with a total of twenty-one pulses with SNR > 5. The single pulse diagnostic plot for this pulsar is shown in the top panel of Figure 6.7. The relatively large number of pulses suggested that this pulsar may also be a periodicity candidate. Results from the Einstein@home pipeline had a weak candidate at the fifth harmonic of the period, but folding at the period and DM determined by the single pulse detection resulted in a solid detection. It is likely that the presence of RFI, combined with the pulsar being relatively weak, resulted in the periodicity search having difficulty properly identifying the pulsar at the correct period. The DM and period determined from the periodicity analysis are 165.7 pc cm$^{-3}$ and 1.727 s.

The single pulse rating statistics for this discovery were $N_{sp}$ = 3.8, $DM_{mean}$ = 162.6pc cm$^{-3}$, and $DM_{std}$ = 1.0. In actuality, ten pulses had SNR > 7. The rating roughly predicted the number of pulses. The estimated rating-based DM is only 2 pc cm$^{-3}$ larger than the value determined by the periodicity analysis.

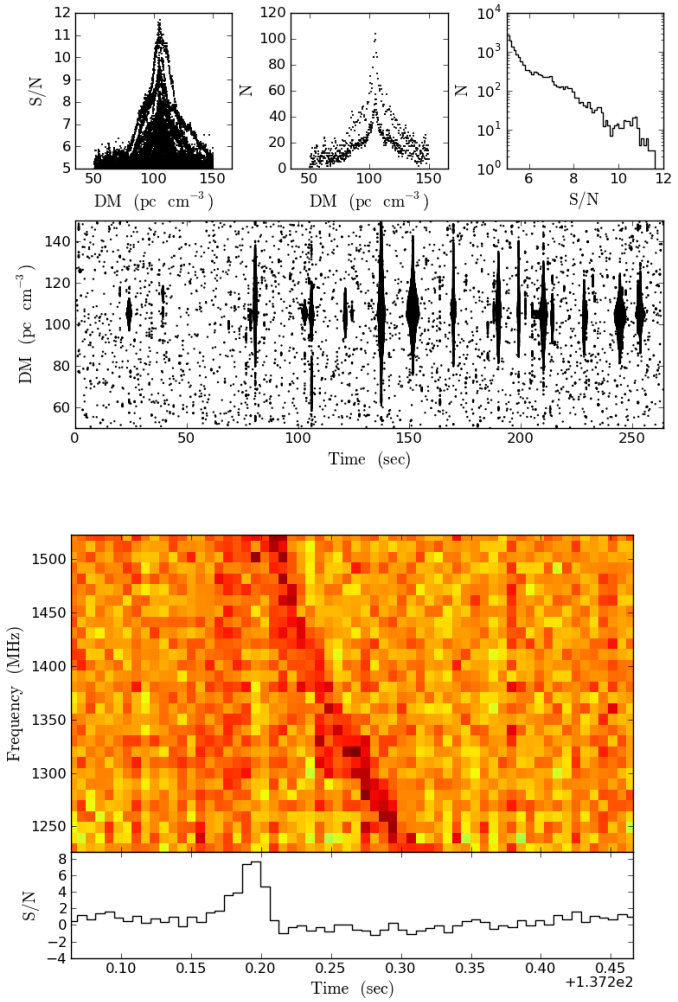


Figure 6.6: Discovery plots of J1928+14. The top four panels show single pulse diagnostic plots for the discovery observations over a narrow range of DMs centered on the true DM of the pulsar. The lower panel shows the time-frequency snapshot of the brightest single pulse.

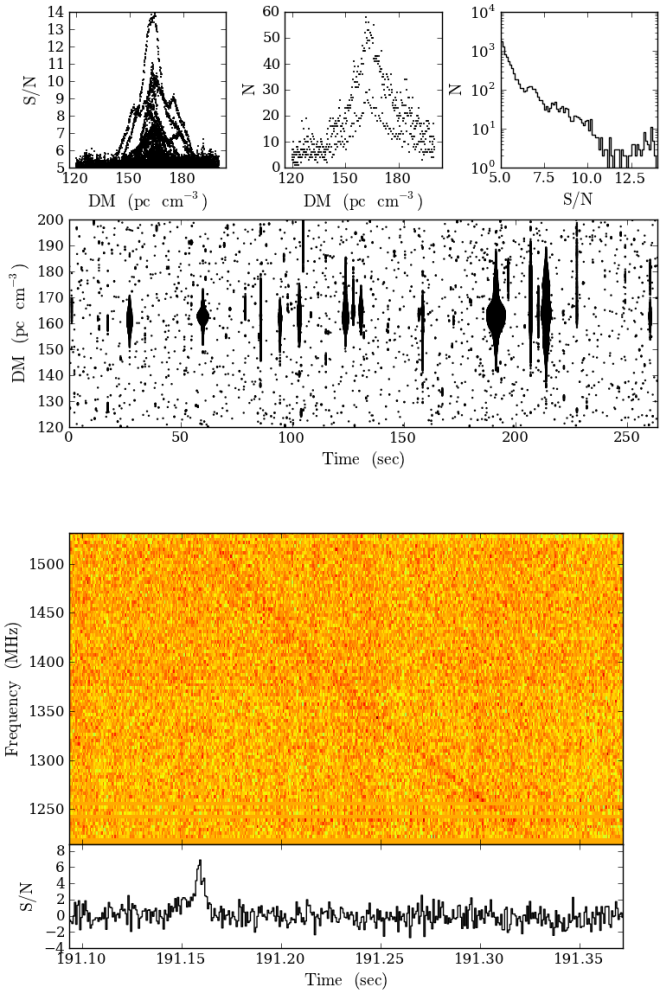


Figure 6.7: Discovery plots of J1905+10. The top four panels show single pulse diagnostic plots for the discovery observations over a narrow range of DMs centered on the true DM of the pulsar. The lower panel shows the time-frequency snapshot of the brightest single pulse at t=191 s.

| Name | RA (hh:mm:ss) | Dec (dd:mm) | $P$ (s) | $W_t$ ($W_{50}$) (ms) | DM (pc cm$^{-3}$) | $S_p$ ($\langle S \rangle$) (mJy) |
|---|---|---|---|---|---|---|
| J1929+11 | 19:29:13(7) | 11:54(2) | 3.217 | 1.3 | 80.5 | 41.7 |
| J1913+06 | 19:13:35(7) | 06:56(2) | 1.257 | 2 | 144.0 | 28 |
| J1928+14 | 19:28:09(7) | 14:44(2) | 1.011 | 2 (47.3) | 104.2 | 35 (0.1) |
| J1905+10 | 19:05:26(7) | 10:33(2) | 1.727 | 3 (33.7) | 165.7 | 34 (0.05) |
| J1917+11 | 19:17:01(7) | 11:42(2) | 5.060 | 3.9 | 319 | 24 |

Table 6.1: Parameters of five new PALFA single pulse pulsar discoveries. The uncertainties on the position correspond to the angular size of the half-power beam width of an individual ALFA beam in the discovery observation. For each pulsar the width ($W_t$) and flux density $S_p$ of the brightest pulse is given. For pulsars that also have periodicity detections, the full width half maximum of the average pulse profile ($W_{50}$) and period-averaged flux density ($\langle S \rangle$) is reported in parenthesis.

### 6.4.5 J1917+11

J1917+11 was discovered in observations conducted on 13 November 2010. Three bright pulses were seen at DM = 319 pc cm$^{-3}$ and are clearly seen in the upper panel of Figure 6.8 at t = 104, 221, and 241 s. The three pulses can be fit to a maximum possible period of 5.06 s. The brightest pulse has SNR = 11 and width of 3.9 ms, yielding a peak flux density of 24 mJy. The time-frequency data for this pulse are shown in the lower panel of Figure 6.8. The single pulse rating statistics for this observations are $N_{sp}$ = 2.0, DM$_{mean}$ = 318.9pc cm$^{-3}$, and DM$_{std}$ = 0.8pc cm$^{-3}$. The rating predicts two pulses when in actuality we observed three pulses with SNR > 7.

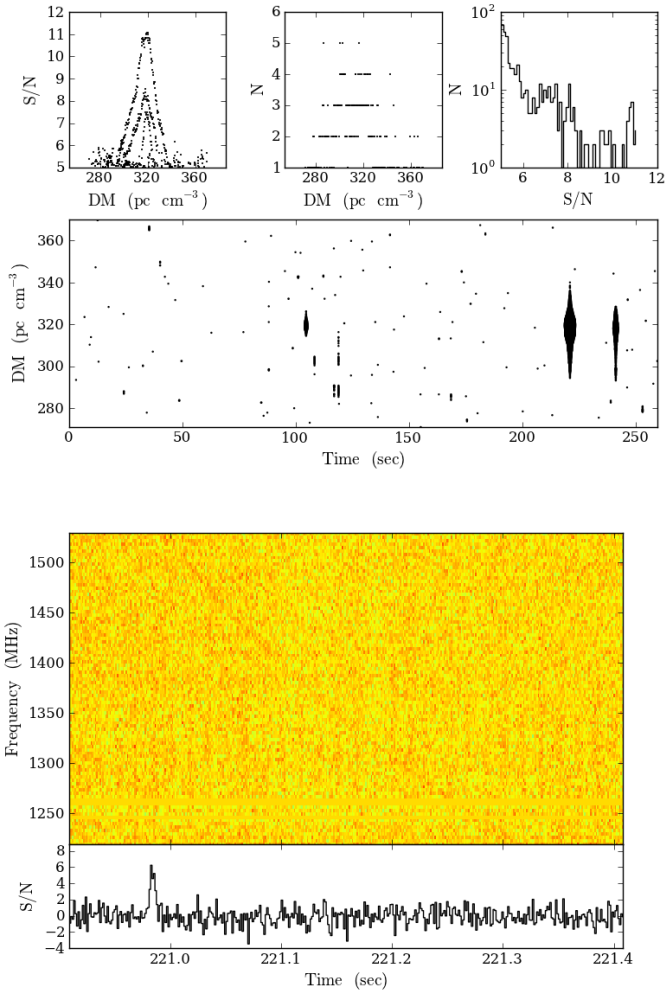


Figure 6.8: Discovery plots of J1917+11. The top four panels show the single pulse diagnostic plots for the discovery observations over a narrow range of DMs centered on the true DM of the pulsar. The bottom panel shows the frequency-time data for the single pulse at t = 221 s.

| Name | $T_{\rm tot}$ (s) | $N_{\rm tot}$ | Rate ($\rm hr^{-1}$) | Comment |
|---|---|---|---|---|
| J1929+11 | 268 | 4 | 54 | § 6.4.1 |
| J1913+06 | 268 | 7 | 94 | § 6.4.2 |
| J1928+14 | 268 | 20 | 269 | § 6.4.3 |
| J1905+10 | 268 | 21 | 282 | § 6.4.4 |
| J1917+11 | 268 | 3 | 40 | § 6.4.5 |

Table 6.2: Observation parameters for the five new PALFA single pulse discoveries.

## 6.5 Analysis

### 6.5.1 Single Pulse Detection Statistics

In this section we outline the number of pulsars discovered using the matched filtering and clustering algorithms. The matched filtering algorithm redetected 47 previously known pulsars in 93 pointings. The cluster algorithm redetected 19 pulsars in 33 pointings, and all 19 were also detected with the matched filter algorithm. The larger number of detections with the matched filtering algorithm is not surprising. The cluster algorithm requires pulses to have a high SNR without any temporal smoothing to be detected, whereas a pulse which is below the SNR threshold in the full resolution time series can be boosted above the noise by the boxcar convolution.

We can estimate the fraction of known pulsars that were detected using single pulse methods. A sample of known pulsars was determined by querying the ATNF catalog[5] (Manchester et al., 2005) for pulsars within 2.625 arcmin of the center of all beams. The search radius of 2.625 arcmin corresponds to 1.5 times

[5] http://www.atnf.csiro.au/people/pulsar/psrcat/

the half-width half-power (HWHP) point. The first null is at roughly twice the HWHP point, so our search radius is half-way between the half power level and the first null. We compared our detections of known pulsars that are within 2.625 arcmin of the beam center and in the ATNF catalog to a list of all pulsars from the ATNF catalog in our pointings. A search of the ATNF catalog yielded 40 unique pulsars in 63 pointings. We redetected 31 unique pulsars in 43 pointings at were both in the ATNF catalog and within the search radius. This suggests $\approx$ 63% of pulsars pointed at can be detected in our single pulse search, and the other 37% have single pulses too weak to detect.

Now we will compare the rates of detection between our processing and the previous single pulse search analysis of PALFA data (Deneva et al., 2009). Deneva et al. (2009) discovered seven new RRATs in a total of 187 deg$^2$, five in the inner galaxy (99 deg$^2$) and two in the outer galaxy (87 deg$^2$). Of the five inner galaxy detections, four were RRATs and one was a canonical pulsar. On average they detected one RRAT per 25 deg$^2$ in the inner galaxy.

In our analysis we found one pulsar per 6 deg$^2$ if we include all five pulsars or one RRAT per 11 deg$^2$ if we exclude the two pulsars that are also visible in periodicity analysis. The volume sampled by the beam varies with bandwidth as $V \propto B^{3/4}$. The Mock spectrometers should therefore sample 2.4 times as much volume per pointing as the WAPP spectrometers if all other observational parameters are kept constant. This calculation is simplistic as it does not take into account more complicated effects such as RFI. Assuming that RRATs are uniformly distributed in distance, scaling the detection rate of 25 deg$^2$ by the appropriate bandwidths yields a predicted detection rate of one pulsar per 10.3 deg$^2$, which is consistent with our observed RRAT detection rate of one per

11 deg$^2$.

Deneva et al. (2009) detected two pulsars in the outer galaxy for a rate of one pulsar per 43 deg$^2$. With the Mock spectrometers we would expect to see one pulsar per 18 deg$^2$ on average. As our analysis only include 9 deg$^2$ of outer galaxy pointings, it is not surprising that we did not discover any new pulsars in these data.

### 6.5.2 Efficacy of the single pulse rating

The single pulse rating and DM statistics as described in Section 6.2 were calculated for all of the beams for all pointings, and the values were displayed along with the seven-panel diagnostic plots, such as that shown in Figure 6.3. In all cases where a pulsar was seen in a beam, the single pulse rating accurately reflected an estimate for the number of pulses with SNR > 7, as well as a good estimate for the DM of the pulsar.

Of our redetections of known pulsars (including multiple detections of the same pulsar), sixty had both a pulse rating (i.e. there were observed pulses with SNR > 7) and a published DM in the ATNF catalog. Only four had an average DM determined from the ratings algorithm that was different from the published DM by more than 10 pc cm$^{-3}$. In one case two pulsars were observed in the same beam, and our simplistic mean and standard deviation metrics could not properly handle this case. The calculated mean was in between the true DM values of the pulsars, and the standard deviation was large. In another case the pulsar was too weak and the rating was actually associated with strong RFI in the beam. In the last case, RFI due to radar was present at similar DMs to the

pulsar, thereby skewing the mean. If these four cases are excluded, the mean and standard deviation of the difference between rating-determined DM and the published DM are 0.3 pc cm$^{-3}$ and 1.8 pc cm$^{-3}$ respectively.

Our experience informs us that the signature of a real pulsar is a pulse rating indicative of one or more dispersed pulses with a non-zero average DM and low DM standard deviation. Feeding this information back into our results, we can ask what, if any, signals pass through this filter that are not pulsars. Searching our results for cases with $DM_{mean} > 10$pc cm$^{-3}$ and $DM_{std} < 5.5$pc cm$^{-3}$ and no pulsar detections, yields around 100 beams (roughly 0.1 % of the total number of beams). Inspecting these beams manually suggests that in the majority of these cases, the ratings were caused by RFI due to airport radars. The airport radar is bright enough to be detectable over a large range DMs, and therefore can look similar to pulsar pulses.

In conclusion, our single pulse rating reliably recognizes real, dispersed pulses. It also misidentifies some broadband RFI as "pulse-like", suggesting this rating needs to be used in conjunction with other methods. In particular an algorithm that better removes RFI from radars would greatly improve the efficacy of the single pulse rating for PALFA data. The interpretation of the pulse ratings beyond the simplistic calculation of the mean and standard deviation of the cluster DM would also improve the rating's usefulness.

### 6.5.3 Pulse flux distribution analysis

Single pulse analysis probes the high-flux tail of the pulse amplitude distribution. The redetections of known pulsars through single pulse analysis allows

one to test models of the pulse amplitude distribution. For each redetection in our data, we will compare the flux density of the pulse having the highest SNR with an estimate of the peak flux density that we should have detected assuming an underlying model for the flux distribution.

The peak flux density of a pulse is defined to be $S$, and we describe the probability density function of the pulse amplitude distribution by $f_S(S)$. The probability of a pulse with $S > S_{\max}$ is

$$F_S(S > S_{\max}) = \int_{S_{\max}}^{\infty} f_S(S) dS. \tag{6.3}$$

Common forms for $f_S(S)$ include log-normal, Gaussian, and power law distributions.

The observed flux density of pulses can also be modulated by diffractive interstellar scintillations (DISS). The effect can be modeled as a gain factor $g_{\rm diss}$, which modifies the unscintillated flux density by $S' = g_{\rm diss} S$ (Cordes & Chernoff, 1997). The probability density function for $g_{\rm diss}$ is given by Cordes & Chernoff (1997) and included here for completeness:

$$f_{g_{\rm diss}}(g_{\rm diss}, n_{\rm DISS}) = \frac{(g_{\rm diss} n_{\rm DISS})^{n_{\rm DISS}}}{g_{\rm diss} \Gamma(n_{\rm DISS})} e^{-g_{\rm diss} n_{\rm DISS}} U(g_{\rm diss}), \tag{6.4}$$

where $U(g_{\rm diss})$ is the Heaviside function, $\Gamma$ is the gamma function, and $n_{\rm DISS}$ is the number of frequency-time scintiles in the data that are averaged in time and frequency. An estimation of the number of scintiles in a time series is given by

$$n_{\rm DISS} \sim \left(1 + 0.2 \frac{B}{\Delta \nu_{\rm d}}\right)\left(1 + 0.2 \frac{T}{\Delta t_{\rm d}}\right) \tag{6.5}$$

where $\Delta \nu_{\rm d}$ and $\Delta t_{\rm d}$ are the characteristic frequency and time scales of the DISS respectively (Cordes, 1986), $B$ is observation bandwidth, and $T$ is the amount of time included in the average. The scintillation bandwidth and time scales are calculated using the NE2001 model (Cordes & Lazio, 2002).

The scintillated flux density probability density function is given by

$$f_{S'}(S') = \int dg_{\rm diss} g_{\rm diss}^{-1} f_{g_{\rm diss}}(g_{\rm diss}, n_{\rm DISS}) f_S(S'/g_{\rm diss}). \tag{6.6}$$

As $n_{\rm DISS} \rightarrow \infty$, the probability density function for $g_{\rm diss}$ tends to a delta function centered at $g_{\rm diss} = 1$, and $f_{S'} \rightarrow f_S$.

For our data $T \sim N_{\rm sm}\Delta t_{\rm s}$ where $\Delta t_{\rm s} = 65.5\,\mu s$ and $N_{\rm sm}$ is the number of samples smoothed, and the typical scintillation timescale for the pulsars in our sample is 10 -100 s. Since $N\Delta t_{\rm s} \ll \Delta t_{\rm d}$, the second factor in Equation 6.5 is equal to 1. The scintillation bandwidths of the pulsars in our sample range from $\sim$ 10 Hz to 100 kHz, compared to a bandwidth of 322 MHz. The number of scintilles being averaged for our sample is $n_{\rm DISS} > 200$, and we can assume $g_{\rm diss} \approx 1$ in all cases.

A log-normal distribution is consistent with flux densities for many pulsars. Burke-Spolaor et al. (2012) analyzed the single pulse "energies" for 315 pulsars observed in the High Time Resolution Universe intermediate-latitude survey and found that one-third of the pulsars were well-matched to a log-normal distribution. It is standard practice to consider the pulse distribution in terms of a normalized flux density $\hat{S} = S/\langle S \rangle$ where $\langle S \rangle$ is the average peak pulse flux density for a given pulsar. The log-normal distribution is given by

$$f_{\hat{S}}(\hat{S}) = \frac{1}{\sqrt{2\pi}\sigma_S \hat{S}} e^{\frac{-(\log\hat{S} - \mu_S)^2}{2\sigma_s^2}} \tag{6.7}$$

where $\sigma_S$ and $\mu_S$ are the scale and location parameter respectively. We adopt the values of $\mu_S = 1.18$ and $\sigma_S = 0.11$ determined by Burke-Spolaor et al. (2012) for pulses with high SNR.

For each pulsar in our sample, we estimate the average flux density, $\langle S_{\rm est} \rangle$, as

$$\langle S_{\rm est} \rangle = S_{\rm cat} \frac{P}{W_{\rm cat}} \tag{6.8}$$

where $P$ is the period of the pulsar and $S_{\rm cat}$ and $W_{\rm cat}$ are catalog values for the period-averaged flux density and pulse width respectively obtained from the ATNF pulsar catalog. When available $S_{1400}$, the average flux density at an observing frequency of 1400 MHz, was used for $S_{\rm cat}$. If there is no entry in the catalog for $S_{1400}$, we used $S_{400}$, the average flux density at an observing frequency of 400 MHz, scaling the flux to 1400 MHz using a spectral index of -1.6[6]. When available, $W_{50}$ was used for $W_{\rm cat}$, otherwise we used $W_{10}$, the full width at 10% maximum. If none of the four observational parameters were available for a pulsar, we excluded it from this analysis. Five pulsars were excluded. There is also uncertainty in the catalog-derived fluxes, and in particular from the pulse width parameters. The values given in the ATNF catalog are derived from observations at different frequencies, and the width of a pulse varies with frequency. The derived catalog flux density is weakly dependent on the pulse width (see Equation 6.10). We estimate that the catalog flux density is uncertain to with a factor of a few.

The maximum flux density we would expect to see in an observation of length $T_{\rm tot}$ seconds is the flux density such that, on average, we see only one pulse. If the total number of pulses in an observation is $N_{\rm p} = T_{\rm tot}/P$, we can write this condition as

$$F_S(S > S_{\rm max,est})N_{\rm p} = 1. \tag{6.9}$$

More generally, the right hand side of this equation is the expected number of pulses with $S > S_{\rm max,est}$. Using the log-normal distribution described by Equation 6.7, we numerically invert the distribution to determine $\hat{S}_{\rm max,est}$ such that the probability of having $\hat{S} > \hat{S}_{\rm max,est}$ is $1/N_{\rm p}$. The estimated peak flux density is then $S_{\rm max,est} = \hat{S}_{\rm max,est}\langle S_{\rm est}\rangle$.

[6]Value is the mean of a distribution of 343 pulsar spectral indices (Lorimer et al., 1995)

The maximum observed flux density ($S_{\rm max,det}$) was calculated by using the SNR of the brightest single pulse and the associated boxcar width converted to a pulse width in ms in each beam that had a redetection of a known pulsar. These were used in the radiometer equation to convert to flux density:

$$S = \mathrm{SNR}\frac{\mathrm{S_{sys}}}{\sqrt{BW_{\rm t}}}\frac{\mathrm{G(0,0)}}{\mathrm{G(B,\phi)}} \tag{6.10}$$

where $S_{\rm sys}$ is the system equivalent flux density of Beam 0. The factor $G(0,0)/G(B,\phi)$ accounts for off-axis detections through a beam-specific gain correction $G(B,\phi)$ where $B$ is the beam index (i.e. 0-6) and $\phi$ is the angular offset from beam center. The gain factor is normalized to the center of Beam 0, such that $G(0,0) = 1$. The gain profiles for each beam were determined from measurements of ALFA's power pattern made by scanning a known, radio galaxy in a grid with 0.25 arcmin spatial resolution. (Cordes et al., 2006). We sliced each beam radially from the center of the beam through the first side lobe to determine $G(B,\phi)$. The offset, $\phi$, is determined by calculating the angular separation between right ascension and declination of the pulsar with the center of the beam it was discovered in. As most of our pulsar detections are off axis, this gain correction is non-negligible.

In Figure 6.9 the maximum detected pulse flux $S_{\rm max,det}$ is shown versus the estimated average flux ($\langle S_{\rm est}\rangle$) for each of the pulsar detections in our sample. The magenta points correspond to detections in the main beam, i.e. only pulsars with positional offset of $\Delta\theta < 2.625$ arcmin from the beam center, which corresponds to an angular separation 1.5 times the half-width at half-max. The cyan points are detections of pulsars in the first side lobe ($4.375 < \Delta\theta < 7$ arcmin).

The solid black line illustrates where $S_{\rm max,det} = \langle S_{\rm est}\rangle$ and is shown for reference. The dashed lines outline the expected region for $S_{\rm max,est}$ assuming a log-

normal distribution as given by Equation 6.7 and corresponds to $\hat{S}_{\mathrm{max,est}} = 2.34$ and 2.77. These values were chosen by taking extreme values for the pulsar period and pulse probability factor ($p_{\mathrm{p}}$). For the lower limit, the shortest period in our sample (36 ms) and a probability factor of 0.1 was used, while for upper limit the longest period in our sample (2.2 s) and a probability factor of 10 was used.

The error bars represent uncertainty in the flux density due radiometer noise and uncertainties in determination of the gain correction factor. A 1-$\sigma$ error ($\sigma_n$) in the flux due to radiometer noise was calculated using Equation 6.10 with SNR = 1 and $G(B, \phi) = 1$. Our gain correction factor was obtained from measurements with limited spatial resolution. Considering only the main beam, we found that the typical gain difference between grid points was $\sim 0.074$. Therefore any detection that falls in a given 0.25 arcmin grid point, has a gain factor which is uncertain to $\pm 0.047$. This error in the gain factor can be converted into an error in flux ($\sigma_g$). The total error is $\sigma_s = \sqrt{\sigma_n^2 + \sigma_g^2}$.

Figure 6.9 includes detections of 28 pulsars in 36 pointings in the main beam and 11 pulsars in 21 pointings in the first side lobes for a total of 39 pulsars in 57 pointings. We include multiple detections of the same pulsar in different beams and pointings as each detection is independent and represents a separate test. Twelve detections (31%) have maximum detected flux densities that are consistent with their predicted $S_{\mathrm{max,est}}$ values. This is consistent with the fraction of pulsars (33%) that Burke-Spolaor et al. (2012) found were well-fitted by a lognormal distribution. All other pulsars lie well above the predicted peak flux for a lognormal distribution suggesting either the amplitudes do not follow a lognormal distribution or that the distribution breaks down at higher $S_{\mathrm{max,est}}$.

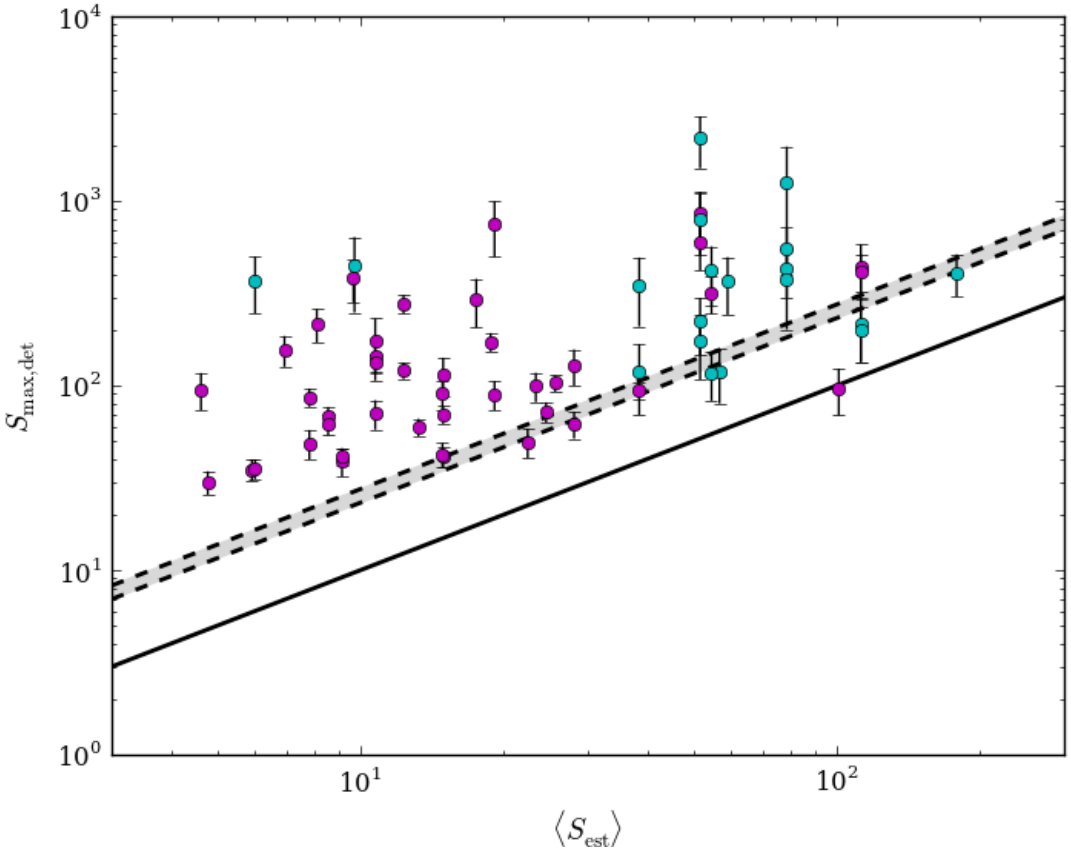


Figure 6.9: Maximum detected pulse flux density ($S_{\mathrm{max,det}}$) vs. estimated average pulse flux density ($\langle S_{\mathrm{est}} \rangle$) for our redetections of known pulses in the main beam (magenta) and first sidelobe (cyan). The solid black line shows where $S_{\mathrm{max,det}} = \langle S_{\mathrm{est}} \rangle$ for reference. The dashed lines delineate the region of predicted peak flux assuming a lognormal pulse amplitude distribution of 2.34 and 2.77 times the estimated average flux. The error bars are the rms of the 1-$\sigma$ flux error from radiometer noise and uncertainties in the gain factor (see Section 6.5.3 for details).

## 6.6 Conclusion

We have presented the results of a single pulse analysis of data from the Pulsar ALFA survey, which includes five new discoveries. Single pulse techniques are crucial in the discovery of highly intermittent pulsars that may not be visible in periodicity searches. The results of single pulse processing could be used as one factor to distinguish a periodicity candidate as real or RFI. Our analysis also shows that many canonical pulsars can also be detected through single pulses with less of a computational burden than is needed for periodicity searches. Pulsar search processing should include both periodicity and single pulse search techniques. Furthermore our discovery of two canonical pulsars that were missed by periodicity pipelines suggests that several different analysis techniques can help mitigate detection difficulties arising from RFI.

These results include several improvements over previous surveys. First, the data were taken with the Mock spectrometers, which have a factor of three improvement in bandwidth over the WAPP spectrometers. The analysis in Section 6.5.1 suggests that our detection rates are higher than in data taken with the WAPP spectrometers by a factor that is commensurate with the increased bandwidth. Our analysis included a new RFI rejection technique that uses the spectral modulation index to remove narrowband RFI and some false positives from noise. Finally, we developed a single pulse rating technique to automatically identify clusters of events that are consistent with the signature of real pulses in the DM versus time plane. Our algorithm reliably identifies real pulses as pulses, but also mis-identifies some RFI as pulses as well.

Future work includes analyzing more data, as we have only analyzed $\approx 30\%$

of the data collected with the Mock spectrometers. Several improvements in RFI rejection could also be made. For example, we are currently not doing any cross-beam anti-coincidence analysis to reject signals that appear in the same region of the DM-time plane in multiple beams. Also, an algorithm that specifically targets the airport radar would clean up the statistical diagnostics and reduce the number of false pulses identified by the single pulse rating algorithms.

# CHAPTER 7

# CONCLUSIONS AND FUTURE WORK

In this dissertation I worked on various aspects of time-domain studies of the radio sky including instrumentation development, signal processing algorithm creation, and the application of these developments to a variety of radio astronomical data.

I helped design, build and deploy a new digital spectrometer used in the SERENDIP V SETI survey operating at the Arecibo Observatory. This project exposed me to all aspects of a large survey including data management, the use of databases, and working in a collaboration. I developed time-domain based algorithms to identify candidate signals from extraterrestrial civilizations and reject radio frequency interference. Data from 2010 were analyzed, and no candidate signals were detected, but we were limited by the few number of sky positions with multiple observations. In the immediate future I want to analyze the rest of the data and expand the algorithms to include frequency drifts in a signal due to Doppler shifts.

We introduce the spectral modulation index and apply it data from two different surveys. It has already proven to be a useful tool for RFI rejection. I hope to expand its use to be a more nuanced tool for characterization and classification and also explore other statistics using even higher order moments. In particular we discuss an alternative definition of the modulation index that may help identify spectra containing residual DM smearing.

The observations of the binary white dwarf system J0651 expanded my time-domain applications to also include long-duration transients and variability.

As an astronomer the experience of observing (setting up the observations and writing an observing script) was invaluable. We have submitted a proposal to conduct follow-up observations of J0651 with Arecibo. We want to redo the 5 GHz observations, this time in a mode that will give us sensitivity to persistent emission, and also observe at 8 GHz to probe higher magnetic fields. At the time of writing, we were still waiting on whether the proposal was given time or not.

The Roof Survey, like the SERENDIP V survey, gave me experience in end to end development of a survey. I was involved in every stage of the survey: the analog receiver chain, building hardware, designing the digital spectrometer, writing RFI rejection algorithms, implementing and expanding a fast transient detection pipeline, and sifting through the results. The data collection and processing ran automatically, controlled by a multi-threaded Python script. While the scientific results may not be profound, the experience that lies outside of scientific publications will be valuable.

Finally, I reprocessed the first 23 months of PALFA data taken with the Mock spectrometers and discovered five new pulsars and redetected 47 previously known pulsars. This project finally gave me the chance to work with actual sources and discover something new. The processing occurred on ATLAS, a large computer cluster at the Max-Planck-Institut-für Gravitationsphysik / Albert-Einstein Institut in Hannover, Germany, and required learning how to interface with large clusters. In the future I plan to continue analyzing Mock data to search for new single pulse candidates.

In conclusion, I have shown that it is valuable, and in some cases crucial, that the radio sky be considered in the time-domain in addition to the frequency

domain.

APPENDIX A

# UPGRADES TO THE ROOFTOP RADIO TELESCOPE

The radio telescope on the roof of the Space Sciences Building underwent a major overhaul in Fall of 2010. The elevation drive in the telescope mount was replaced. The analog signal path electronics were upgraded and a new digital spectrometer was developed to replace the existing hardware.

## A.1 System Overview

The telescope is a 3.8 m parabolic reflector on an AZ-EL mount facing south. The telescope can point over an elevation range of 0 to 90 deg and an azimuth range of 160 to 240 degrees (180 being due south). It is equipped with a helical, single polarization L-band feed at the prime focus. The feed has a wide bandwidth, so it is limited to 1-1.5 GHz with a bandpass filter before being amplified by a low noise amplifier. After the LNA the signal travels down to a rack in the lab on the sixth floor for further conditioning.

A detailed diagram of the stages in the analog conditioning can be seen in Figure A.1. (Everything described in the previous paragraph is contained in the "ANT" block in this diagram.) The RF signal is split and sent through two bandpass filters. One of these filters selects a 57 MHz wide band centered on 1420 MHz for observations of the 21-cm line of neutral hydrogen. The second selects a band centered on 1667 MHz for observations of the OH line. A computer-controlled analog RF switch allows the observer to choose which filter to observe with. The signal is then mixed with the first local oscillator that brings

the RF band down to an intermediate frequency with a center frequency of 260 MHz. The frequency of the local oscillator is computer controlled and can be tuned from 1 - 1.5 GHz. After the first mixing stage, the signal goes through an anti-aliasing filter with a center frequency of 260 MHz and bandwidth of 100 MHz. The new spectrometer samples the IF 1 directly, but an additional mixing stage is done to maintain backwards compatibility. A second local oscillator mixes the signal at 210 MHz resulting in a band centered on 50 MHz (IF 2) and another anti-aliasing bandpass filter is used. This signal is sampled by the old hardware.

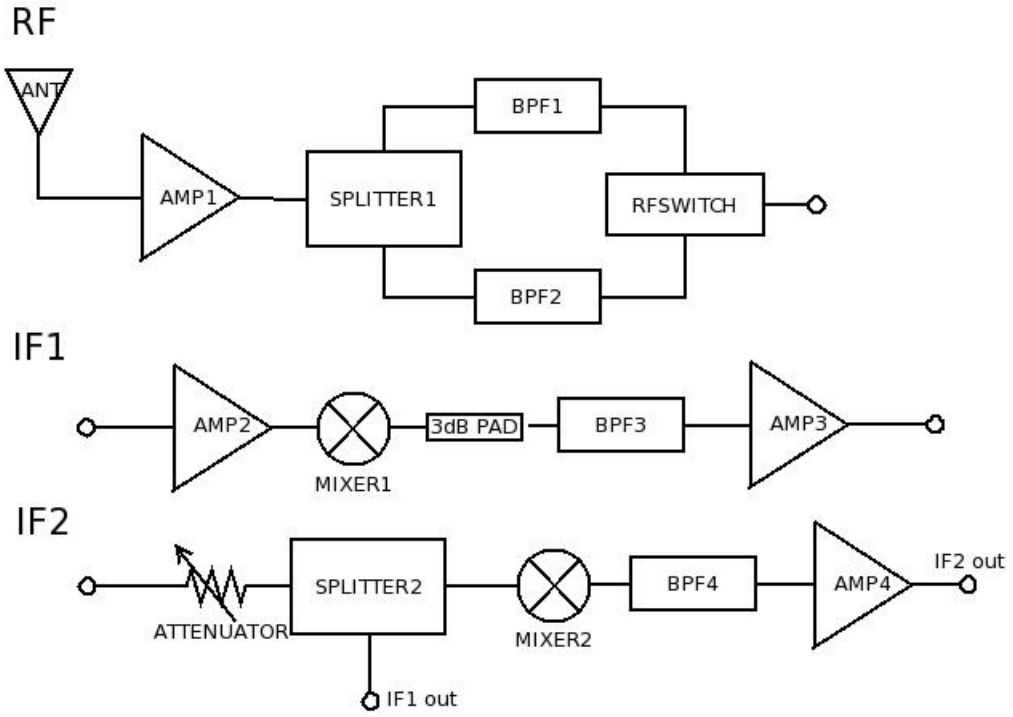


Figure A.1: Diagram of analog signal chain after system upgrades.

The following describes the computer controlled components in more detail.

### A.1.1 Computer controlled analog components

#### TTL board

A D-type flip flop drives two TTL signals to BNC jacks. The data pin and clock pin for each flipflop are connected to pins D0-D3 on the DB-25 connector. Software sets either Hi or LO on the data pin and then latches it in. See Figure2 for diagram. This board must be attached to one of the parallel ports connected directly to the PCI slot (parport1 or parport2), not an extender. See software section for usage.

#### toggle_ttl.c

Sets TTL1 or TTL2 HI and LO.
Usage: ./a.out TTL line TTLstate
Example 1: ./a.out 1 1
Sets TTL line 1 HI
Example 2: ./a.out 2 0
Sets TTL line 2 LO

#### set_dsattenuator.c

Sets the attenuation level of the digital step attenuator. Valid values are 0 to 31.5 dB in increments of 0.5 dB.
Usage: ./a.out attnlevel
Example: ./a.out 23.5 will set the attenuation to 23.5 dB

**ns_lo1.c**

Sets the frequency of the 1000-1500 MHz Novasource, which is used as the first local oscillator. This code has been minimally modified from the DOS code we received from Nova Engineering. Some DOS-specific functions were changed to their Linux equivalent. Also there were some rounding issues in converting the floating point number given by the user to integers. The current code is accurate to 1 MHz.
Usage: ./a.out lo_freq
Example: ./a.out 1420.0 will set the novasource to 1420 MHz

### A.1.2 Cables

The SMA cables are RG-316 double shielded coax of lengths 6 and 36 inches.

### A.1.3 Novasource configuration

Black boxes used as clocks and local oscillator sources. The frequency and attenuation settings can be set using the Windows-based GUI that is on the Novasource CD. In addition the frequency of the RF output can be set using C code based on a DOS program obtained from Nova Engineering and modified by Laura. Currently it's only ever been tested with the 1000-1500 MHz box but would in principle be made to work with the other two. The only way to adjust the attenuation presently is using the Windows GUI. Each box has a stored configuration that it loads upon power up and are listed in the table below. Note:

The attenuation level ranges from 0 - 31 where zero corresponds to around +13 dBm and 31 to -21 dBm as measured by Laura on the power meter.

| Model Num | Freq range | Freq setting | Atten setting | Output |
|---|---|---|---|---|
| NS2-1000104 | 1000 - 1500 MHz | 1160 MHz | 8 | +5 dBm |
| NS2-0160104 | 160 - 220 MHz | 210 MHz | 8 | +5 dBm |
| NS2-0080503 | 80 - 120 MHz | 90 MHz | 11 | +3 dBm |

Table A.1: Novasource stored parameters

## A.2 iBOB Rooftop spectrometer

The iBOB+iADC[1] based spectrometer for use with the telescope on the roof is a basic digital FFT spectrometer. The 260 MHz IF is real sampled by the iADC at a clock rate of 416 Msamples/sec (see Figure A.2). The 208-416 MHz Nyquist band is aliased down to 208-0 MHz (i.e. the band is flipped in frequency). The iBOB complex mixes this signal at 156 MHz, applies a 50 MHz FIR bandpass filter, and decimates to a 100 MHz bandwidth. A 8192-point complex FFT is calculated. The complex FFT channels are converted to a 32-bit unsigned power and accumulated to a user-definable length in a vector accumulator. The resulting accumulated spectra are packetized in the PowerPC and sent to a PC over UDP. A single UDP packet contains a 16-byte header and 256 channels (assuming 32bits per channel).

[1] `http://casper.berkeley.edu`

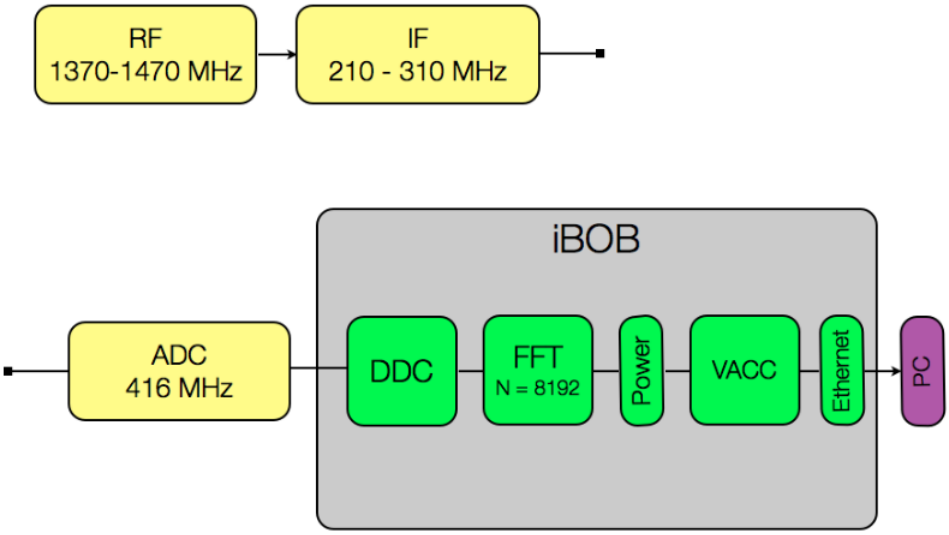


Figure A.2: Overview of RF/IF path and iBOB block diagram

### A.2.1 Firmware

The firmware (or "personality") of the FPGA is designed with the graphic design tool *Simulink* that is part of Matlab. The CASPER libraries were used in order to target the iBOB board.

### A.2.2 Software

Data is transferred from the iBOB to a PC using UDP over the 10/100 Mb port. Although this it is simple enough to write C code to capture UDP packets, I've chosen to use the network utility 'gulp' because it is highly optimized for large data rates. Likely this is overkill, but it is easy to use and writes to the standard 'pcap' format. Also note that gulp needs to be run as root. The raw pcap files can then be processed by a modified version of sigproc to be converted to filterbank

format, averaged, header read, etc.

### A.2.3 Software for a4410

Several scripts and executables are used to interface the iBOB spectrometer with the two standard a410 data collection scripts */scripts/chart_recorder_rspec* and */scripts/spectrum_analyzer6*. Details of their use are provided in DOC2BWRITTEN.

**Added to ~/bin**

start_rspec - Two arguments: dwell_time num_points. Kills any existing gulp process and starts new gulp process. Must be 'sudo'ed so a410 is given permission to run it as root.
stop_rspec - Kills any running gulp process. Must be 'sudo'ed so a410 is given permission to run it as root.
load_mod - Checks to see if appropriate parallel port modules are loaded correctly and loads them if not. Needs to be 'sudo'ed. Run by iBOB configuration scripts.
ch_pcapfile - Changes permission on pcap files generated by gulp so non-root users can access them. Needs to be 'sudo'ed.
gulp
filterbank_rspec - Converts .pcap file to filterbank (.fil) formatted file.
bandpass - sigproc program that averages a filterbank file over time. The default is to average the entire file into one spectrum. Note the scripts may reference the same program in /../lspitler/mysvn/sigproc-4.3/.

**Added to ~/script**

chart_recorder_rspec - Chart recorder script modified to use ibob roof spec
chart_recorder.sm - Modified (not added) to include cr3 macro for roof spec
spectrum_analyzer6 - Spectrum analyzer version that uses roof spec
spectrum_analyzer6.sm - SM script modified for roof spec
rspec_get_spec - Primary script for getting spectra/power from iBOB
starting_HI_spec - Programs the iBOB's personality and configures the software registers. Writes status to "starting_ibob.log" and make use of "load_ibob.imp".

### A.2.4 Example spectra

Figures A.3 and A.4 show data from a drift scans using the roof telescope system. Figure A.3 illustrates how the bandpass shape can be removed. See the Figure caption for details. Figure A.4 shows a long drift scan and the drift in the frequency of HI as the Earth rotates.

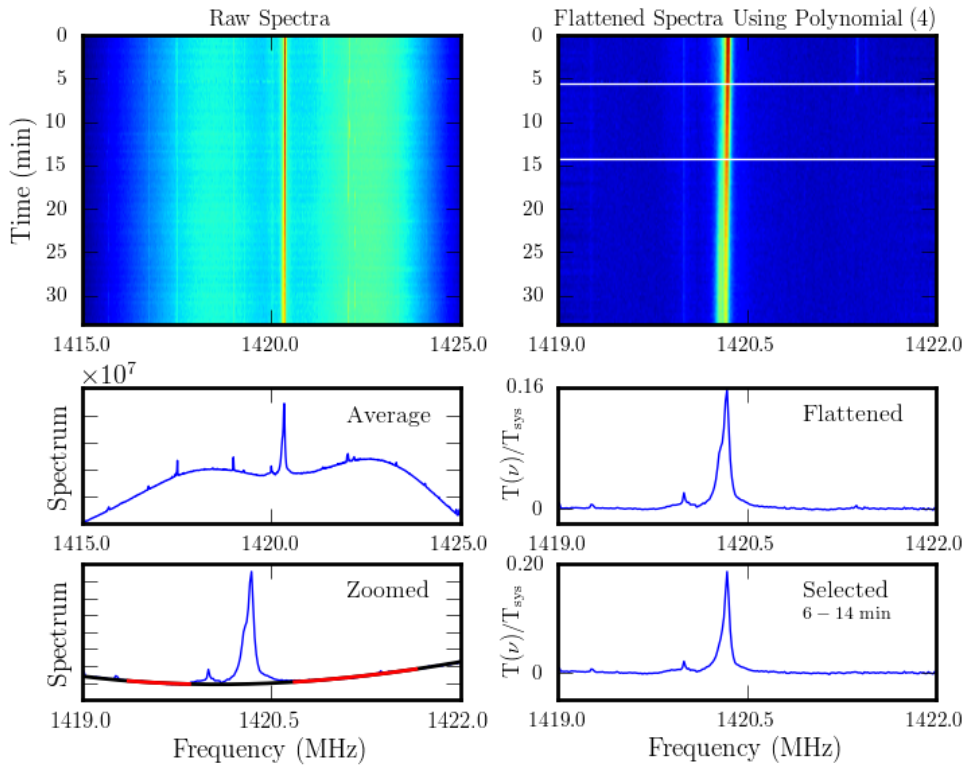


Figure A.3: Short duration drift scan using the roof telescope system. The top two frames show dynamic spectra of 40 min. The panel on the left is the raw data, and the panel on the right is the same data with the baseline subtracted. The two smaller panels on the left show an average spectrum across the entire bandpass (top), and a region focused on the neutral hydrogen line and an over plot of the polynomical fit (bottom). The two smaller panels on the right show a bandpass subtracted spectra in units of the system temperature for the entire duration (top) and a selected region of time (bottom) indicated with the white horizontal lines in the right dynamic spectrum.

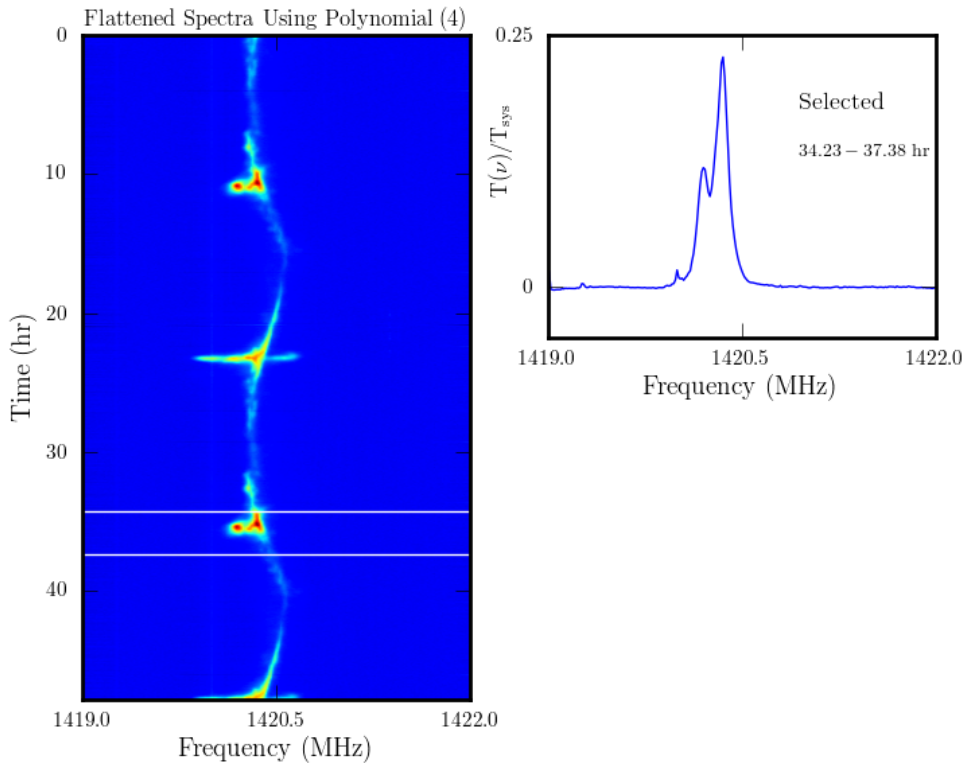


Figure A.4: Example data taken with the roof telescope system. The left panel shows a bandpass subtracted dynamic spectrum of a 48-hour drift scan. The signal in the center of the plot drifting in frequency is neutral hydrogen. The horizontal white lines indicate the region of time used to make the spectrum shown on the right.

## BIBLIOGRAPHY

Aznar Cuadrado, R., Jordan, S., Napiwotzki, R., Schmid, H. M., Solanki, S. K., & Mathys, G. 2004, A&A, 423, 1081

Backer, D. C. 1970, Nature, 228, 42

Backus, P. R., & Project Phoenix Team. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 278, Single-Dish Radio Astronomy: Techniques and Applications, ed. S. Stanimirovic, D. Altschuler, P. Goldsmith, & C. Salter, 525–527

Bastian, T. S., Bookbinder, J., Dulk, G. A., & Davis, M. 1990, ApJ, 353, 265

Berger, E. 2002, ApJ, 572, 503

Bhat, N. D. R., Cordes, J. M., Cox, P. J., Deneva, J. S., Hankins, T. H., Lazio, T. J. W., & McLaughlin, M. A. 2011, ApJ, 732, 14

Blandford, R. D. 1977, MNRAS, 181, 489

Booth, R. S., de Blok, W. J. G., Jonas, J. L., & Fanaroff, B. 2009, ArXiv e-prints

Bowyer, S., Korpela, E., Werthimer, D., Cobb, J., Lampton, M., Lampton, M., & Lebofsky, M. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 213, Bioastronomy 99, ed. G. Lemarchand & K. Meech, 473

Bowyer, S., Werthimer, D., & Donnelly, C. 1995, in Astronomical Society of the Pacific Conference Series, Vol. 74, Progress in the Search for Extraterrestrial Life., ed. G. S. Shostak, 285

Bowyer, S., Werthimer, D., Donnelly, C., Lampton, M., Herrick, W., Soller, J., Ng, D., & Hiatt, T. 1993, in Astronomical Society of the Pacific Conference Se-

ries, Vol. 47, Third Decennial US-USSR Conference on SETI, ed. G. S. Shostak, 269

Bowyer, S., Zeitlin, G., Tarter, J., Lampton, M., & Welch, W. J. 1983, Icarus, 53, 147

Brown, W. R., Kilic, M., Allende Prieto, C., & Kenyon, S. J. 2010, ApJ, 723, 1072

—. 2012, ApJ, 744, 142

Brown, W. R., Kilic, M., Hermes, J. J., Allende Prieto, C., Kenyon, S. J., & Winget, D. E. 2011, ApJL, 737, L23

Burke-Spolaor, S., & Bailes, M. 2010, MNRAS, 402, 855

Burke-Spolaor, S., Bailes, M., Ekers, R., Macquart, J.-P., & Crawford, III, F. 2011a, ApJ, 727, 18

Burke-Spolaor, S., Bailes, M., Johnston, S., Bates, S. D., Bhat, N. D. R., Burgay, M., D'Amico, N., Jameson, A., Keith, M. J., Kramer, M., Levin, L., Milia, S., Possenti, A., Stappers, B., & van Straten, W. 2011b, MNRAS, 416, 2465

Burke-Spolaor, S., Johnston, S., Bailes, M., Bates, S. D., Bhat, N. D. R., Burgay, M., Champion, D. J., D'Amico, N., Keith, M. J., Kramer, M., Levin, L., Milia, S., Possenti, A., Stappers, B., & van Straten, W. 2012, MNRAS, 423, 1351

Cai, Y.-F., Sabancilar, E., Steer, D. A., & Vachaspati, T. 2012, PRD, 86, 043521

Camilo, F., Ransom, S. M., Chatterjee, S., Johnston, S., & Demorest, P. 2012, ApJ, 746, 63

Cobb, J., Lebofsky, M., Werthimer, D., Bowyer, S., & Lampton, M. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 213, Bioastronomy 99, ed. G. Lemarchand & K. Meech, 485

Cocconi, G., & Morrison, P. 1959, Nature, 184, 844

Cordes, J. M. 1976, ApJ, 210, 780

—. 1986, ApJ, 311, 183

—. 1997, Square Kilometer Array Memo 97, http://www.skatelescope.org

Cordes, J. M., Bhat, N. D. R., Hankins, T. H., McLaughlin, M. A., & Kern, J. 2004, ApJ, 612, 375

Cordes, J. M., & Chernoff, D. F. 1997, ApJ, 482, 971

Cordes, J. M., Freire, P. C. C., Lorimer, D. R., Camilo, F., Champion, D. J., Nice, D. J., Ramachandran, R., Hessels, J. W. T., Vlemmings, W., van Leeuwen, J., Ransom, S. M., Bhat, N. D. R., Arzoumanian, Z., McLaughlin, M. A., Kaspi, V. M., Kasian, L., Deneva, J. S., Reid, B., Chatterjee, S., Han, J. L., Backer, D. C., Stairs, I. H., Deshpande, A. A., & Faucher-Giguère, C.-A. 2006, ApJ, 637, 446

Cordes, J. M., & Lazio, T. J. 1991, ApJ, 376, 123

Cordes, J. M., & Lazio, T. J. W. 2002, ArXiv Astrophysics e-prints

Cordes, J. M., Lazio, T. J. W., & Sagan, C. 1997, ApJ, 487, 782

Cordes, J. M., & McLaughlin, M. A. 2003, ApJ, 596, 1142

Cordes, J. M., & Rickett, B. J. 1998, ApJ, 507, 846

Cropper, M., Harrop-Allin, M. K., Mason, K. O., Mittaz, J. P. D., Potter, S. B., & Ramsay, G. 1998, MNRAS, 293, L57

Dall'Osso, S., Israel, G. L., & Stella, L. 2007, A&A, 464, 417

de Vos, M., Gunst, A. W., & Nijboer, R. 2009, IEEE Proceedings, 97, 1431

Deller, A. 2010, in RFI Mitigation Workshop

Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., Nice, D., Ransom, S., Stairs, I. H., Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., Chamberlin, S. J., Cordes, J. M., Ellis, J., Finn, L. S., Freire, P., Giampanis, S., Jenet, F., Kaspi, V. M., Lazio, J., Lommen, A. N., McLaughlin, M., Palliyaguru, N., Perrodin, D., Shannon, R. M., Siemens, X., Stinebring, D., Swiggum, J., & Zhu, W. W. 2012, ArXiv e-prints

Demorest, P. B., Pennucci, T., Ransom, S. M., Roberts, M. S. E., & Hessels, J. W. T. 2010, Nature, 467, 1081

Deneva, J. S., Cordes, J. M., McLaughlin, M. A., Nice, D. J., Lorimer, D. R., Crawford, F., Bhat, N. D. R., Camilo, F., Champion, D. J., Freire, P. C. C., Edel, S., Kondratiev, V. I., Hessels, J. W. T., Jenet, F. A., Kasian, L., Kaspi, V. M., Kramer, M., Lazarus, P., Ransom, S. M., Stairs, I. H., Stappers, B. W., van Leeuwen, J., Brazier, A., Venkataraman, A., Zollweg, J. A., & Bogdanov, S. 2009, ApJ, 703, 2259

Drake, F. D. 1961, Physics Today, 14, 40

Eatough, R. P., Molkenthin, N., Kramer, M., Noutsos, A., Keith, M. J., Stappers, B. W., & Lyne, A. G. 2010, MNRAS, 407, 2443

Edwards, R. T., Bailes, M., van Straten, W., & Britton, M. C. 2001, MNRAS, 326, 358

Farrell, W. M., Desch, M. D., & Zarka, P. 1999, JGR, 104, 14025

Fuller, J., & Lai, D. 2012, MNRAS, 421, 426

Gajjar, V., Joshi, B. C., & Kramer, M. 2012, MNRAS, 424, 1197

Gaudi, B. S. 2012, ARAA, 50, 411

Goldreich, P., & Lynden-Bell, D. 1969, ApJ, 156, 59

Grießmeier, J.-M., Zarka, P., & Spreeuw, H. 2007, A&A, 475, 359

Hallinan, G., Antonova, A., Doyle, J. G., Bourke, S., Brisken, W. F., & Golden, A. 2006, ApJ, 653, 690

Hankins, T. H., & Rickett, B. J. 1975, in Methods in Computational Physics. Volume 14 - Radio astronomy, ed. B. Alder, S. Fernbach, & M. Rotenberg, Vol. 14, 55–129

Hansen, B. M. S., & Lyutikov, M. 2001, MNRAS, 322, 695

Hermes, J. J., Kilic, M., Brown, W. R., Winget, D. E., Allende Prieto, C., Gianninas, A., Mukadam, A. S., Cabrera-Lavers, A., & Kenyon, S. J. 2012, ArXiv e-prints

Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, Nature, 217, 709

Hulse, R. A., & Taylor, J. H. 1975, ApJL, 195, L51

Israel, G. L., Panzera, M. R., Campana, S., Lazzati, D., Covino, S., Tagliaferri, G., & Stella, L. 1999, A&A, 349, L1

Jackson, P. D., Kundu, M. R., & White, S. M. 1989, A&A, 210, 284

Jacoby, B. A., Bailes, M., Ord, S. M., Edwards, R. T., & Kulkarni, S. R. 2009, ApJ, 699, 2009

Johnston, S., Feain, I. J., & Gupta, N. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 407, The Low-Frequency Radio Universe, ed. D. J. Saikia, D. A. Green, Y. Gupta, & T. Venturi, 446–+

Kalogera, V., Kim, C., Lorimer, D. R., Burgay, M., D'Amico, N., Possenti, A., Manchester, R. N., Lyne, A. G., Joshi, B. C., McLaughlin, M. A., Kramer, M., Sarkissian, J. M., & Camilo, F. 2004, ApJL, 614, L137

Keane, E. F., Ludovici, D. A., Eatough, R. P., Kramer, M., Lyne, A. G., McLaughlin, M. A., & Stappers, B. W. 2010, MNRAS, 401, 1057

Keane, E. F., & McLaughlin, M. A. 2011, Bulletin of the Astronomical Society of India, 39, 333

Keane, E. F., Stappers, B. W., Kramer, M., & Lyne, A. G. 2012, MNRAS, 425, L71

Kedziora-Chudczer, L. L., Jauncey, D. L., Wieringa, M. H., Tzioumis, A. K., & Reynolds, J. E. 2001, MNRAS, 325, 1411

Kilic, M., Brown, W. R., Allende Prieto, C., Agüeros, M. A., Heinke, C., & Kenyon, S. J. 2011, ApJ, 727, 3

Knight, H. S. 2006, Chinese Journal of Astronomy and Astrophysics Supplement, 6, 020000

Korpela, E. J. 2012, Annual Review of Earth and Planetary Sciences, 40, 69

Korpela, E. J., Cobb, J., Fulton, S., Lebofsky, M., Heien, E., Person, E., Demorest, P., Bankay, R., Anderson, D., & Werthimer, D. 2004, in IAU Symposium, Vol. 213, Bioastronomy 2002: Life Among the Stars, ed. R. Norris & F. Stootman, 419

Kramer, M., Lyne, A. G., O'Brien, J. T., Jordan, C. A., & Lorimer, D. R. 2006, Science, 312, 549

Lai, D. 2012, ArXiv e-prints

Landstreet, J. D., Bagnulo, S., Valyavin, G. G., Fossati, L., Jordan, S., Monin, D., & Wade, G. A. 2012, A&A, 545, A30

Lonsdale, C. J., Cappallo, R. J., Morales, M. F., Briggs, F. H., Benkevitch, L., Bowman, J. D., Bunton, J. D., Burns, S., Corey, B. E., Desouza, L., Doeleman, S. S., Derome, M., Deshpande, A., Gopala, M. R., Greenhill, L. J., Herne, D. E., Hewitt, J. N., Kamini, P. A., Kasper, J. C., Kincaid, B. B., Kocz, J., Kowald, E., Kratzenberg, E., Kumar, D., Lynch, M. J., Madhavi, S., Matejek, M., Mitchell, D. A., Morgan, E., Oberoi, D., Ord, S., Pathikulangara, J., Prabu, T., Rogers, A., Roshi, A., Salah, J. E., Sault, R. J., Shankar, N. U., Srivani, K. S., Stevens, J., Tingay, S., Vaccarella, A., Waterson, M., Wayth, R. B., Webster, R. L., Whitney, A. R., Williams, A., & Williams, C. 2009, IEEE Proceedings, 97, 1497

Lorimer, D. R. 2008, Living Reviews in Relativity, 11, 8

Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, Science, 318, 777

Lorimer, D. R., & Kramer, M. 2004, Handbook of Pulsar Astronomy, ed. R. Ellis, J. Huchra, S. Kahn, G. Rieke, & P. B. Stetson

Lorimer, D. R., Yates, J. A., Lyne, A. G., & Gould, D. M. 1995, MNRAS, 273, 411

Lyne, A. G., & Graham-Smith, F. 2006, Pulsar Astronomy

Magro, A., Karastergiou, A., Salvini, S., Mort, B., Dulwich, F., & Zarb Adami, K. 2011, ArXiv e-prints

Manchester, R. N., Fan, G., Lyne, A. G., Kaspi, V. M., & Crawford, F. 2006, ApJ, 649, 235

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993

Manchester, R. N., Lyne, A. G., Camilo, F., Bell, J. F., Kaspi, V. M., D'Amico, N., McKay, N. P. F., Crawford, F., Stairs, I. H., Possenti, A., Kramer, M., & Sheppard, D. C. 2001, MNRAS, 328, 17

McLaughlin, M. A., & Cordes, J. M. 2003, ApJ, 596, 982

McLaughlin, M. A., Lyne, A. G., Lorimer, D. R., Kramer, M., Faulkner, A. J., Manchester, R. N., Cordes, J. M., Camilo, F., Possenti, A., Stairs, I. H., Hobbs, G., D'Amico, N., Burgay, M., & O'Brien, J. T. 2006, Nature, 439, 817

Montebugnoli, S., Cattani, A., Cecchi, M., Maccaferri, A., Monari, J.-M., Mariotti, S., Cosmivici, C. B., & Maccone, C. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 213, Bioastronomy 99, ed. G. Lemarchand & K. Meech, 501

Murphy, T., Chatterjee, S., Kaplan, D. L., Banyer, J., Bell, M. E., Bignall, H. E., Bower, G. C., Cameron, R., Coward, D. M., Cordes, J. M., Croft, S., Curran, J. R., Djorgovski, S. G., Farrell, S. A., Frail, D. A., Gaensler, B. M., Galloway, D. K., Gendre, B., Green, A. J., Hancock, P. J., Johnston, S., Kamble, A., Law, C. J., Lazio, T. J. W., Lo, K. K., Macquart, J.-P., Rea, N., Rebbapragada, U., Reynolds, C., Ryder, S. D., Schmidt, B., Soria, R., Stairs, I. H., Tingay, S. J., Torkelsson, U., Wagstaff, K., Walker, M., Wayth, R. B., & Williams, P. K. G. 2012, ArXiv e-prints

Nice, D. J. 1999, ApJ, 513, 927

Nita, G. M., Gary, D. E., Liu, Z., Hurford, G. J., & White, S. M. 2007, PASP, 119, 805

Piddington, J. H., & Drake, J. F. 1968, Nature, 217, 935

Pshirkov, M. S., & Postnov, K. A. 2010, Ap&SS, 330, 13

Ramsay, G., Brocksopp, C., Wu, K., Slee, B., & Saxton, C. J. 2007, MNRAS, 382, 461

Ramsay, G., Cropper, M., Wu, K., Mason, K. O., & Hakala, P. 2000, MNRAS, 311, 75

Ramsay, G., Hakala, P., & Cropper, M. 2002, MNRAS, 332, L7

Rees, M. J. 1977, Nature, 266, 333

Rickett, B. J., Hankins, T. H., & Cordes, J. M. 1975, ApJ, 201, 425

Rohlfs, K., & Wilson, T. L. 2004, Tools of radio astronomy

Ruf, C., Renno, N. O., Kok, J. F., Bandelier, E., Sander, M. J., Gross, S., Skjerve, L., & Cantor, B. 2009, Geophys. Res. Lett., 361, L13202

Siemion, A. P. V., Bower, G. C., Foster, G., McMahon, P. L., Wagner, M. I., Werthimer, D., Backer, D., Cordes, J., & van Leeuwen, J. 2012, ApJ, 744, 109

Siemion, A. P. V., Cobb, J., Chen, H., Cordes, J., Filiba, T., Foster, G., Fries, A., Howard, A., von Korff, J., Korpela, E., Lebofsky, M., McMahon, P. L., Parsons, A., Spitler, L., Wagner, M., & Werthimer, D. 2011, ArXiv e-prints

Spangler, S. R., & Spitler, L. G. 2004, Physics of Plasmas, 11, 1969

Spitler, L. G., Cordes, J. M., Chatterjee, S., & Stone, J. 2012, ApJ, 748, 73

Staveley-Smith, L., Wilson, W. E., Bird, T. S., Disney, M. J., Ekers, R. D., Freeman, K. C., Haynes, R. F., Sinclair, M. W., Vaile, R. A., Webster, R. L., & Wright, A. E. 1996, PASA, 13, 243

Stootman, F. H., De Horta, A. Y., Oliver, C. A., & Wellington, K. J. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 213, Bioastronomy 99, ed. G. Lemarchand & K. Meech, 491

Sullivan, III, W. T., Brown, S., & Wetherill, C. 1978, Science, 199, 377

Thompson, D. R., Wagstaff, K. L., Brisken, W. F., Deller, A. T., Majid, W. A., Tingay, S. J., & Wayth, R. B. 2011, ApJ, 735, 98

Vachaspati, T. 2008, Physical Review Letters, 101, 141301

van Haasteren, R., Levin, Y., Janssen, G. H., Lazaridis, K., Kramer, M., Stappers, B. W., Desvignes, G., Purver, M. B., Lyne, A. G., Ferdman, R. D., Jessner, A., Cognard, I., Theureau, G., D'Amico, N., Possenti, A., Burgay, M., Corongiu, A., Hessels, J. W. T., Smits, R., & Verbiest, J. P. W. 2011, MNRAS, 414, 3117

Wang, N., Manchester, R. N., & Johnston, S. 2007, MNRAS, 377, 1383

Wayth, R. B., Tingay, S. J., Deller, A. T., Brisken, W. F., Thompson, D. R., Wagstaff, K. L., & Majid, W. A. 2012, ApJL, 753, L36

Weisberg, J. M., Armstrong, B. K., Backus, P. R., Cordes, J. M., Boriakoff, V., & Ferguson, D. C. 1986, AJ, 92, 621

Weisberg, J. M., & Taylor, J. H. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 328, Binary Radio Pulsars, ed. F. A. Rasio & I. H. Stairs, 25

Werthimer, D., Bowyer, S., Cobb, J., Lebofsky, M., & Lampton, M. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 213, Bioastronomy 99, ed. G. Lemarchand & K. Meech, 479

Willes, A. J., & Wu, K. 2004, MNRAS, 348, 285

Wolszczan, A., & Frail, D. A. 1992, Nature, 355, 145

Wu, K., Cropper, M., Ramsay, G., & Sekiguchi, K. 2002, MNRAS, 331, 221

Yardley, D. R. B., Coles, W. A., Hobbs, G. B., Verbiest, J. P. W., Manchester, R. N., van Straten, W., Jenet, F. A., Bailes, M., Bhat, N. D. R., Burke-Spolaor, S., Champion, D. J., Hotan, A. W., Oslowski, S., Reynolds, J. E., & Sarkissian, J. M. 2011, MNRAS, 414, 1777

Zarka, P. 1998a, JGR, 103, 20159

—. 1998b, JGR, 103, 20159

Zarka, P., Farrell, W. M., Kaiser, M. L., Blanc, E., & Kurth, W. S. 2004, Planet. Space Sci., 52, 1435